



THE LIBRARY
OF
THE UNIVERSITY
OF CALIFORNIA
LOS ANGELES



Digitized by the Internet Archive
in 2007 with funding from
Microsoft Corporation

BOOKS BY H. L. HOLLINGWORTH.

The Inaccuracy of Movement,—*Archives of Psychology*, No. 13, (Columbia Contributions to Philosophy and Psychology, Vol. XVII, No. 3) pp. 87. June, 1909. New York. The Science Press. 80 cents.

The Influence of Caffein on Mental and Motor Efficiency,—*Archives of Psychology*, No. 22, (Columbia Contributions to Philosophy and Psychology, Vol. XX, No. 4) pp. 167. April, 1912. New York. The Science Press. \$1.50 (paper), \$1.75 (cloth).

Principles of Appeal and Response,—(A Systematic Textbook of Business Psychology) pp. 315. New York. 1913. D. Appleton and Company. \$2.00 net. By mail, \$2.16.

Experimental Studies in Judgment,—*Archives of Psychology*, No. 29, (Columbia Contributions to Philosophy and Psychology, Vol. XXII, No. 3). pp. 125. December, 1913. New York. The Science Press. \$1.25 (paper), \$1.50 (cloth).

EXPERIMENTAL STUDIES IN JUDGMENT

H. L. HOLLINGWORTH
COLUMBIA UNIVERSITY

*Columbia Univ.
Psychology R R*

ARCHIVES OF PSYCHOLOGY

EDITED BY
R. S. WOODWORTH

No. 29, DECEMBER, 1913

COLUMBIA CONTRIBUTIONS TO PHILOSOPHY AND PSYCHOLOGY,
VOL. XXII, NO. 3

NEW YORK
THE SCIENCE PRESS

AGENTS G. E. STECHERT & CO.; London (2 Star Yard, Carey St., W. C.); Leipzig (Hospital St., 10);
Paris (76, rue de Rennes).

MEMORANDUM
FOR THE RECORD
DATE

Page 50
11-345.8

II 153.5

4 725

13C
181

14725e

TABLE OF CONTENTS

	PAGE
INTRODUCTION	V
CHAPT. I. Judgments of Personal Efficiency	1
II. Perceptual Criteria of Judgments of Efficiency..	17
III. Performer and Witness as Judges of Efficiency..	27
IV. The Central Tendency of Judgment.....	44
V. The Direction of Judgment.....	53
VI. Natural or Habitual Tendencies of Judgment...	59
VII. Judgments of Similarity and Difference.....	68
VIII. Influence of Form and Category on the Outcome of Judgment	85
IX. The Perceptual Basis for Judgments of Extent of Movement	92
X. Some Characteristics of Judgments of Evaluation	96

1260715

INTRODUCTION

A GENERAL title, such as that given to this monograph, can give very little preliminary indication of the nature of the problems therein suggested or investigated. In the study of those mental processes, acts or resultants which we vaguely call *judgments* there are perhaps four chief problems with which special researches may be concerned:

(a) The *nature* and *mechanism* of judgments. Studies which have sought for introspective ear-marks or criteria of the judgment process,—qualitative differentia between judgments and other elementary or complex states or processes or acts, belong here. Here also would belong any attempt to describe or hypothecate the physiological correlate of judgments. With these problems the studies here presented are not concerned.

(b) The *forms*, *varieties* and *classification* of judgments. This may be conceived as a task for logical rather than for psychological inquiry. It may suffice here merely to indicate that these studies are in no primary way concerned with problems of classification.

(c) The *basis* or *perceptual criteria* of typical judgments,—the data which determine the content, direction, or outcome of special varieties of judgments under given conditions. Two of the studies here presented are specifically directed toward this type of problem. Thus in Chapter II. and in Chapter IX. attempts are made to discover on what data one relies when he judges the efficiency of a work process or the extent or duration of a voluntary movement.

(d) The *laws* or *behavior* of judgments, and the ways in which the laws are modified or the behavior conditioned by specific variations of the judgment situation. Among these specific variations of the judgment situation may be mentioned, by way of examples, the form in which the judgment is expressed, the category employed, the nature of the material to be judged, individual, age, sex and group differences, previous practise, preceding judgments, habitual judgment tendencies, etc. On problems of this sort all of the studies here presented have more or less direct bearing.

The studies have been made from a fairly definite point of view, or at least they have been actuated by a fairly permanent interest. Stated in general terms, this has been an interest in *the way in which mind works* rather than in *what is in the mind* at the moment of its operation. As I have elsewhere remarked, such an interest finds but

little use for the introspective method. It is an interest "not in the momentary content of a conscious moment; nor in the descriptive character of the sensory fragment which may at that moment be the bearer of meaning; nor in the instrument, criterion or vehicle of an act of apprehension, a comparison, a feeling, or a choice." It is above all an interest in "the outcome of this moment in the form of behavior,—an act, a choice, a judgment, and in the character, reliability, constancy, and significance which the outcome of such a mental operation possesses."

Of the ten studies which the volume contains, six are entirely new and have not been elsewhere reported. The remaining four have already appeared in the psychological periodicals. They are reprinted here because of their relevance to the later studies and because they were originally part of the larger plan of which this monograph is a partial result.

EXPERIMENTAL STUDIES IN JUDGMENT

CHAPTER I

JUDGMENTS OF PERSONAL EFFICIENCY

INVESTIGATORS of fatigue have frequently found occasion for the remark that the individual's judgments of the quality of his own performance in a piece of work in progress or just completed are far from being a reliable index either of the capacity of his organism at the time, or of the actual amount, speed, or quality of the work done. The matter usually rests, however, with this generalization. No attempts seem to have been made to determine experimentally the reliability of such judgments, except in the cases of a few studies of the confidence of simple sensory discriminations. In a sense, of course, the task of judging the intensity, extent, or duration of two sensory impressions may be called work, even though no emphasis be laid on the number of such judgments to be made in a given unit of time. But sensory discrimination is not to be called work in the active sense indicated in such processes as the production of ergograms, the execution of tapping movements at maximal speed, or the similar high speed performances of "naming opposites," "naming colors," or mental calculation.

In this chapter will be reported a preliminary attempt to investigate the characteristics, conditions, tendencies, and reliability of a worker's judgments of the efficiency of his own performance in such active processes as those just mentioned. Such questions as the following will define the nature of the problem, indicate the direction taken by the present inquiry, and suggest the importance of the topic to that sort of psychology which is interested in the dynamic aspects of the life of psycho-physical organisms.

1. How reliably can a performer judge the quality of his own performance when no objective measures are at his disposal? To what extent is the conscious concomitant of an action a guarantee of the quality or effects of that action?

2. What are the criteria which constitute the basis of one's judgments of his own efficiency at a given moment, or through a given period of time?

3. What are the conditions which modify the character and accuracy of such judgments, both in the same task and in the case of

10102
REVILL
611

different tasks? How do the characteristics of the judgment of personal efficiency change with the conditions of variation and with the nature of the performance?

4. What relations exist between the certainty or degree of confidence of such judgments and their accuracy as shown by objective record?

5. How do the judgments of the performer compare in these respects with the judgments of a witness who observes the progress of the work without participating in it, and without knowledge of the objective records?

6. Do practise, fatigue, transfer, and similar processes affect the course and reliability of these judgments?

7. What individual differences exist in these various respects? How does proficiency in performance correlate with reliability of judgment?

Such questions as these open up a large field of inquiry which has hardly been explored in even a preliminary way. The present study is limited to perhaps three of these problems, and must even here be considered as hardly more than suggestive. It will achieve its main purpose if it succeeds in directing attention toward the general field in which it lies. Further problems of a similar kind will be taken up in Chapters II. and III.

Several investigators, interested mainly in the determination of the differential threshold, in the examination of the psycho-physical relations and methods in the field of sensation, and in the measurement of recognition memory, have taken occasion to instruct their observers to state, in the case of each judgment of sensory discrimination, recognition, etc., the degree of confidence with which the judgment was expressed. Since the present study constitutes the application of a similar procedure to judgments of the efficiency of performance in a work process, a brief account of the most important results of these studies may well be given here.

Fullerton and Cattell¹ while investigating the perception of small differences in extent and speed of movement, lifted weights, and intensity of lights, proceeded mainly by the methods of right and wrong cases and average error. But these methods were combined with the method of just observable differences by requesting the observer to state, after each judgment of difference, the degree of his confidence in his judgment. Three degrees of confidence were used,—*A*, *B*, and *C*, indicating, respectively, “quite confident,” “fairly confident,” and “less confident.” Among the conclusions based on these results the following are of special interest in the present connection:

¹ Fullerton and Cattell, “On the Perception of Small Differences.”

Extent of Movement.—" . . . with regard to the degrees of confidence *a*, *b*, and *c*, it may be objected that the terms 'quite confident,' 'fairly confident,' and 'less confident' are extremely vague. In a series of experiments with the one observer each of these terms may be assumed, perhaps, to have approximately the same meaning in different parts of the series; but the quantitative relations of the subjective feeling of confidence in the three cases remain very obscure, nor can it be assumed that they may be measured by the percentage of right cases corresponding to each degree of confidence. The fact that an observer is always right when he feels quite confident, and right 97 per cent. of the time when he feels fairly confident, does not prove that the amount or degree of his confidence in the two instances is as 100 to 97" (p. 63).

Weights.—"The confidence (of *A* and *B* judgments) varies nearly as the percentage of right cases (with varying sense differences) and some reliance may therefore be placed on such introspection. We see however . . . that different individuals place very different meanings on the degree of confidence. . . . Those observers who felt the greatest degree of confidence in their judgment had the largest probable error, while those who were least seldom quite confident had the smallest probable error. . . . We see that an observer is more apt to be right than wrong, even when he feels very little confidence in the correctness of his decision. We also obtain a rough measure of what reliance may be placed on the judgment of the observer" (p. 126).

Lights.—"The confidence of the observer is hence a fair measure of the correctness of his judgment, but it is evident that *A* and *B* have a widely different meaning in the case of the several observers. . . . It is worth noting that when the discrimination was equally good the confidence was less with lights than with weights" (p. 144).

Griffing's² observers, in judging sensations of pressure and impact, also estimated their degree of confidence in each judgment in some experiments. Griffing concludes, on this point: "The degree of confidence in the perception of intensive differences varies greatly for individuals, the proportion of wrong judgments of which observers were confident ranging from 1/3 to 1/50. The probability of correctness was for most observers from .8 to .9. There is no relation between either of these quantities and the accuracy of discrimination. The percentage of correct guesses (*D* judgments) varied from 52 per cent. to 70 per cent., the average being 59 per cent."

Henmon,³ in a study the chief object of which was the correlation

² Griffing, "On Sensations from Pressure and Impact," *Psych. Mon.*, Vol. I., No. 1.

³ Henmon, "Time and Accuracy of Judgment," *Psych. Rev.*, May, 1911.

of the speed with the accuracy of judgments of visual linear magnitudes, also instructed his observers to assign their degree of confidence to each judgment. He used four degrees of certainty, designated as "perfectly confident," "fairly confident," "with little confidence," and "doubtful." Henmon's chief conclusions on this aspect of his problem are as follows:

"The time of judgment increases uniformly as the degree of confidence decreases. The time of wrong judgments is on the average longer than that of right judgments, while under each category the wrong judgments are in general shorter. The time of wrong judgments is more variable than that of right, and there are indications of two kinds of wrong judgments,—those too quick and those prolonged beyond a certain optimal time. The degree of confidence varies, from subjects who are perfectly confident in 90 per cent. of 500 judgments to those who are perfectly confident in less than 10 per cent. While there is a positive correlation on the whole between accuracy and degree of confidence, the latter is not a reliable index of the former. Subjects whose judgments are quick are neither more nor less accurate than those whose judgments are slow."

In experiments on the effect of length of series on recognition memory, Strong instructed his subjects to grade the confidence of their recognitions of pages of advertisements. Three degrees of certainty were used,—"absolutely certain," "reasonably sure," and "very doubtful." Pure guesses were not required. So far as his conclusions bear on the subject of the present study they are as follows:

"The accuracy approximates 0 with 'very doubtful' recognitions, regardless of the length of the series. . . . Recognitions not accompanied by a feeling of absolute certainty are practically no better than random guesses. . . . As the difficulty of the task increases, the ratio of 'absolutely certain' recognitions to 'reasonably sure' and 'doubtful' recognitions decreases." In general, "we have approximately three fourths the accuracy in pile No. 2 ('reasonably sure') that we find in pile No. 1 ('absolutely certain') and one half the accuracy in pile No. 3 ('doubtful') that we find in pile No. 1." These results were found only when the various observers and the various tasks were combined. "It was not the case with the individual subjects. . . . With each successive series, implying a difference in the difficulty of the task, the relationship between the three piles changed."⁴ In a later study Strong has also investigated the degree of confidence of recognitions of words, after varying intervals.

⁴ Strong, "Effect of Length of Series on Recognition Memory," *Psych. Rev.*, Nov., 1912.

The Present Experiments

In order to secure an adequate situation for the study of judgments of personal efficiency in an active work process, four features must be provided for:

1. The task should be one in which the performer has reached a practise level of performance which closely approximates his physiological or psychological limit. Work on this level of performance will show variations in both directions from an average degree of proficiency. These variations in the directions of "better" and "worse" performance will be approximately equal, except that occasional large inferior records may be made, thus producing variations which can not be equalled by deviations in the direction of "better." It is possible that because of this fact, the ideal place for such work would be on the secondary slope of the practise curve. But there should at any rate be no considerable excess of superior performances such as would occur if the worker were still on the primary slope of the curve of practise.

2. The conditions of performance and the technique of record should be such that, although objective measure of the work is secured, the performer shall have no direct knowledge of these data. The judgment should be based solely on his introspective impressions of the ease, smoothness, agreeableness, or speed of his work. For this situation to be attained the most successful plan is to keep the amount and quality of the work constant and to make the speed of performance (recorded by a second person) the objective measure of efficiency.

3. Various types of tasks should be examined, ranging from work which is chiefly motor and fairly automatic to work which is mainly mental in character. An intermediate stage should also be represented, and is afforded by tests involving perceptual reactions. In the motor work the observer will be enabled to attend more or less directly and objectively to the progress of the work, on the perceptual level more or less attention will be demanded by the details of the process, and observation will be less direct. In the more exclusively mental work attention may be supposed to be quite occupied with the immediate details of performance, and the judgment will be still less direct in character. It is quite conceivable that as one passes from stage to stage the criteria of the judgment of efficiency will shift from one ground to another or others. The introspective analysis of these criteria constitutes a profitable direction of inquiry.

4. The various tasks, to be strictly comparable, should be about equally difficult, should continue for about the same time, should be equally practised, and should yield about the same per cent. of correct judgments.

As tasks which satisfy the above requirements and which are at the same time technically convenient and fairly well standardized, the following three well-known laboratory tests were chosen.

Stage 1. The Tapping Test.—Performer, holding short stylus in right hand, elbow resting on table, tapped 400 times on metal plate at maximal speed. Each tap was recorded by an electric counter and the total time taken with the stop-watch.

Stage 2.—The Color-naming Test.—The Woodworth-Wells blank was used, the colors being named in the same order at each trial. The test blank shows 100 patches of color, each 1 cm. square, and separated by spaces of 1 cm. from its neighbors. Each of the five colors blue, red, green, black, and yellow, is repeated twice in each of the 10 lines of 10 colors each. All sequences of the same color are avoided, as are frequent occurrences of the same sequence of colors. The colors are to be named in order, as in reading, as rapidly as possible. The total time was taken with the stop-watch. No errors were permitted.

Stage 3. Naming Opposites of Words.—A series of 50 adjectives used by the writer in a previous study. The performer was required to go down the list, giving in turn the opposite (antonym) of each word and to complete the list as quickly as possible. The total time was recorded with the stop-watch. At each successive trial the order of occurrence of the words was changed, each order being a chance one. No errors were permitted.

Each test was repeated daily during the major part of the experiment. During the later days two daily trials were made. In order to eliminate practise effect, 60 trials of each test were made (covering a period of two months) before the feature of the experiment here reported was introduced. By this time all the performers (three in number) had practically reached a practise level and during the succeeding 72 trials, on which the present study is based, the average amount of gain in the three tests was but slight. The only exception is the color-naming test, which allowed a certain amount of memory. The average records at the beginning of the practise curve, after the 60 preliminary trials, and at the close of the experiment, were as follows:

Test	Initial Average	Average after 60 Trials	Average at Close of Experiment
Tapping.....	45.5 sec.	39.0 sec.	38.0 sec.
Color-naming.....	44.0	37.0	28.0
Naming opposites.....	46.7	29.0	26.0

The three tests seem to satisfy to a sufficient degree the conditions just enumerated as requisite. Each observer, after each trial in each task, judged his performance to have been either "better than usual"

or "worse than usual," and assigned a degree of confidence to his judgment. Four degrees of confidence were used,—*A* (absolutely certain), *B* (fairly certain), *C* (slightly certain), and *D* (a mere guess). All records were kept from the performer's knowledge and no computations were made, on the point under investigation, until the experiment was completed. One of the observers (*H*) was the writer. Of the other two (*G* and *L*) *G* was a college undergraduate music student, with no psychological training. *L* was a graduate student, with psychological training and with considerable experience both as subject and as experimenter.

The experiment thus required 132 trials in each of three tasks, by each of three observers, a total of 1,188 trials. The first 60 trials in each test were used for the two purposes of reaching practise level and of giving some sort of definition to the term "as well as usual." The remaining trials (648 in all) were used for the judgments of personal efficiency. In computing results, the median of the 7 trials preceding the trial being judged was taken as the standard of comparison. The term "as usual" was found to refer no further back than the previous half dozen days or trials. The median was chosen rather than the average because it makes due allowance for occasional large variations, which the introspections of the observers showed to be allowed for in the judgments of performance. Each trial is thus compared with the median of the 7 trials immediately preceding it. The direction and amount of difference between the two serve as the objective measure of the efficiency of the trial in question. Comparison of this measure with the observer's subjective estimate of his performance will in this way afford a measure of the correctness of his judgment. Comparison of the amount of this difference with the degree of confidence will show the relation of the feeling of certainty to the variation in performance. Since the time of the performance is not quite the same in all tests nor for all observers (although very nearly so in both cases) in some of the tables the absolute differences between standard and single trial are converted into percentages of the total time for the individual or task in question.

Table I. gives the average results for the three observers, for each of the three processes, showing the average deviation from the usual performance on which each degree of confidence was based (A.S.D. = Average Stimulus Difference). The first part of the table gives the absolute variations in seconds, the latter part giving these variations when expressed as per cent. of the average total time required for the test in question. Table II. gives the same results, when assembled regardless of sign or direction of variation, but classified according

to degree of confidence only. Table III. gives the per cent. correctness of all these degrees of confidence and in both directions of variation. This table also gives the distribution of these judgments, thus showing the number of cases on which each average is based. Table IV. gives these same records, regardless of sign. Table V. gives the total distribution of the judgments when classified merely as "judgments of better" and "judgments of worse." The table also gives the actual distribution of the records when thus classified. Both absolute numbers and percentages are given. The sign — is used to indicate "better" (requiring less time) and + to indicate "worse" (requiring longer time) than usual.

TABLE I

SHOWING ABSOLUTE AND PERCENTILE DEVIATIONS FROM "USUAL" ON WHICH THE VARIOUS DEGREES OF CONFIDENCE WERE BASED; CALLED, IN FOLLOW-PAGES, A.S.D (AVERAGE STIMULUS DIFFERENCE). TABLE GIVES AVERAGE CONSTANT ERRORS AND AVERAGE M.V.'s FROM THESE CONSTANT ERRORS

Test	Better			
	A A.S.D. M.V.	B A.S.D. M.V.	C A.S.D. M.V.	D A.S.D. M.V.
Seconds: Tapping. .	- 1.5 0.8	-1.2 0.9	-0.7 0.8	-0.5 1.3
Colors. . . .	- 2.9 1.3	-1.5 1.5	-0.6 1.3	-0.8 1.3
Opposites. .	- 3.0 0.8	-1.7 1.3	-0.7 1.2	-0.1 1.4
Per cent.: Tapping. .	- 3.9	-3.2	-1.7	-1.3
Colors. . . .	- 9.6	-5.0	-2.0	-2.8
Opposites. .	-11.6	-6.6	-2.7	-0.4
Av. per cent.	- 8.4	-4.9	-2.1	-1.5

Test	Worse			
	A A.S.D. M.V.	B A.S.D. M.V.	C A.S.D. M.V.	D A.S.D. M.V.
Seconds: Tapping. .	+ 2.5 0.9	+1.4 1.0	+0.9 0.6	+0.9 1.0
Colors. . . .	+ 3.7 0.6	+2.0 1.2	+1.6 1.3	+0.6 1.9
Opposites. .	+ 5.4 1.2	+1.6 1.2	+1.3 1.7	+0.7 1.4
Per cent.: Tapping. .	+ 6.5	+3.6	+2.4	+2.4
Colors. . . .	+12.3	+6.6	+5.1	+2.0
Opposites. .	+21.0	+6.0	+5.2	+2.6
Av. per cent.	+13.3	+5.4	+4.2	+2.3

TABLE II

SHOWING STIMULUS DIFFERENCES REGARDLESS OF THEIR DIRECTION

Tests	Absolute Differences				Percentile Differences			
	A	B	C	D	A	B	C	D
Tapping.	2.0	1.3	.8	.7	5.2	3.4	2.0	1.9
Color-naming.	3.3	2.0	1.1	.7	10.9	5.8	3.6	2.4
Opposites.	4.2	1.6	1.0	.6	16.3	6.3	3.9	1.5
Averages.	3.2	1.6	1.0	.7	10.8	5.2	3.2	1.9

TABLE III

SHOWING THE CORRECTNESS AND DISTRIBUTION OF THE VARIOUS DEGREES OF CONFIDENCE

Test		Better				Worse			
		-A	-B	-C	-D	+A	+B	+C	+D
Per cent. correct	Tapping.....	87	82	74	57	100	77	79	72
	Color-naming....	100	82	58	59	100	80	79	58
	Opposites.....	100	85	69	53	100	78	70	56
	Averages.....	96	83	67	56	100	79	76	62
Distribution of the Judgments	Tapping.....	31	37	36	27	15	19	24	27
	Color-naming....	16	38	43	43	9	15	19	33
	Opposites.....	12	42	40	27	15	15	28	37
	Totals.....	59	117	119	97	39	49	71	97

TABLE IV

SHOWING CORRECTNESS AND DISTRIBUTION OF THE JUDGMENTS REGARDLESS OF SIGN

Test	Percentile Correctness				Distribution			
	A	B	C	D	A	B	C	D
Tapping.....	94	80	77	65	46	56	60	54
Color-naming.....	100	81	68	59	25	53	62	76
Opposites.....	100	81	70	54	27	57	68	64
Averages.....	98	81	73	59	Totals 98	166	190	194

TABLE V

SHOWING THE DISTRIBUTION OF THE JUDGMENTS AND OF THE ACTUAL RECORDS, WITH RESPECT TO "BETTER" AND "WORSE"

	Test	Worse	Better	Total
Distribution of the Judgments	Tapping.....	85 (39%)	131 (61%)	216
	Color-naming.....	76 (35%)	140 (65%)	216
	Opposites.....	95 (44%)	121 (56%)	216
	Totals.....	256 (39%)	392 (61%)	648
Distribution of the Actual Cases	Tapping.....	99 (46%)	117 (54%)	216
	Color-naming.....	94 (43%)	122 (57%)	216
	Opposites.....	92 (42%)	124 (58%)	216
	Totals.....	285 (44%)	363 (56%)	648

Several interesting points are suggested by these tables:

1. The observer's judgments of the efficiency of his own performance, in successive daily trials in these tests, have a reliability which varies with the confidence of the judgments. Judgments of "absolutely certain" are always correct (100 per cent.) except in the case of judgments of superior performance in tapping, where the average per cent. correctness of the three observers is 87 per cent. Judgments which are "fairly certain" and "slightly certain" show 80 per cent. and 70 per cent. correctness respectively. "Pure guesses"

are correct in 60 per cent. of the cases. In all tests with all observers the correctness of pure guesses is greater than that to be expected from mere chance. This result accords with those of earlier investigations on judgments of sensory discrimination (Cattell, Griffing, Henmon, Jastrow, etc.).

2. Judgments of "better" seem to be based on smaller variations than are judgments of "worse." If the average of all three tests is regarded this is true of all degrees of confidence. Almost twice as great per cent. inferiority is found for a given type of judgment of "worse" as that per cent. of superiority required to produce a judgment of "better." Considering the tests separately this rule holds of all the judgments except in the cases of the *B* judgments in opposites and the *D* judgments in color-naming, in which cases no considerable difference whatever is present. In the case of the three observers this rule holds without exception in the case of the *A* judgments in all tests. The remaining degrees of confidence do not show the relation clearly in the individual records.

There are three possible explanations of this apparently finer discrimination in the case of judgments of superior efficiency.

A.—It may indicate merely a predisposition on the part of the performer to judge his work as good rather than as poor, thus revealing only a prejudice in favor of judgments of "better." If this is the case, the variations in performance on which these "better" judgments are based will be small because of the frequent occurrence of inferior trials which are judged to be superior. This would result in a reduction of the threshold for the class of judgments in question, since frequent + variations would cancel the larger — variations. But if this were the case the judgments of "better" would show a lower percentage of correctness than that of the judgments of "worse" since the latter would have been based for the most part on only the more pronounced cases of inferior performance.

But reference to the table which gives the correctness of the various classes of judgments does not clearly show this to have been the case. In the case of opposites the "better" judgments are no less correct than are the judgments of "worse." In fact the total correctness is slightly higher in the case of the former. In color-naming the same thing is true for *A*, *B*, and *D* judgments. Only in the case of the *C* judgments is there an exception. Tapping alone affords a slightly greater percentage of correctness in the case of the "worse" judgments. The average results of the three tests give 76 per cent. and 79 per cent. correct in the two directions. Or if the categories be disregarded in the computation of correctness, 75 per cent. of the "worse" judgments are correct and 76 per cent. of the "better."

The judgments of "better" are then about as correct on the whole as those of "worse," and this in spite of the fact that the former are based on much smaller variations in efficiency. It does not yet seem then that prejudice in favor of efficiency judgments affords adequate explanation of the differences in threshold.

B.—The relation may be supposed to follow from the mere fact that, when a performer is approximating his physiological level there will occur very few large deviations in the direction of superiority, whereas occasional lapses, interferences, distractions, and accidents might produce large deviations in the direction of inferiority. These large deviations then would tend to increase the average variations from the standard in the case of the judgments of "worse" beyond the point which might be actually necessary as the ground for the given type of judgment.

The possibility that the larger variations for "worse" judgments are merely the result of accidental large inferior deviations is not so

TABLE VI

SHOWING THE DISTRIBUTION OF THE ACTUAL RECORDS (DEVIATIONS FROM
"USUAL"), WITH RESPECT TO THEIR MAGNITUDE

	Sec.	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9
Tapping:	<i>H</i> +	9	15	7	1					
	—	17	8	11	1					
	<i>S</i> +	6	12	8	3	0	1	1		
	—	10	13	12	3	1	0	0		
	<i>L</i> +	14	13	3	1					
	—	25	14	2	0					
	Total +	29	30	18	5	0	1	1	0	0
	—	52	35	25	4	1	0	0	0	0
Color naming:	<i>H</i> +	9	6	4	5	2	3	3	1	
	—	9	6	7	8	4	2	1	0	
	<i>S</i> +	10	8	4	4	0	1			
	—	12	14	5	5	2	1			
	<i>L</i> +	5	6	2	3	2	2		1	
	—	20	10	9	5	3	0		0	
	Total +	24	20	19	12	4	6	3	2	0
	—	41	30	21	18	9	3	1	0	0
Opposites:	<i>H</i> +	7	7	8	4	4	2	1		1
	—	13	12	8	1	3	1	0		0
	<i>S</i> +	11	2	8	2	3	3	0		
	—	14	11	9	6	2	0	1		
	<i>L</i> +	6	9	5	4	2	0	1		1
	—	18	12	6	4	1	0	0		0
	Total +	24	18	21	10	9	5	2	0	2
	—	45	35	23	11	6	1	1	0	0

easily disposed of, but there seems to be sufficient evidence to show that this factor is not the only one at work. As a matter of fact, when the + and — variations are grouped, as in Table VI., according to their magnitude, there is found to be no excessive number of inferior records, although such were theoretically possible, and would perhaps have occurred had not the performers been both zealous and competitive, and nearly on a practise level. In tapping, the largest variations (over 3 sec.) show almost equal distribution for all subjects. The — variations predominate in the smaller groups as the result of slight practise in the course of the experiment. In color-naming the variations are larger than in tapping, but since there continued to be considerable practise in this test, large — deviations are just as frequent as are large inferior records. In fact, with *G* the former are more numerous. But the color-naming shows the superiority of judgments of “better” for *A*, *B*, and *C* degrees of confidence, and one need not expect to find it in the *D* judgments, which were pure guesses. In the case of opposites we clearly have a preponderance of large inferior trials, with all observers. If this is the factor which is responsible for the higher averages of the “worse” judgments, we ought then to find this result most striking in the opposites test. But just the reverse is the case. Opposites is just the test which affords several exceptions to the generalization.

Moreover, even if there were a considerable excess of large positive deviations (worse) these would only affect necessarily the *A* judgments. The *B*, *C*, and *D* judgments would still be based on variations chosen by the observer at will. But the *B*, *C*, and *D* judgments show the same tendency, on the whole, as do the *A* judgments,—smaller variations for judgments of “better” than for equally confident judgments of “worse.”

C.—The present indication seems to be, then, that efficiency is judged on the basis of smaller variations than is inefficiency. Does this mean that the criteria of judgments of efficiency are more definite or more numerous or more clearly detected, and hence that the “feeling of efficiency” arises on smaller provocation than does the “feeling of inefficiency”? The point constitutes an interesting problem for future work, and will be taken up again in a later chapter.

3. Progressively larger variations in performance (both absolute and relative) are required as the basis of judgments of a given degree of confidence, as one passes from tapping, through color-naming, to opposites. With *A* judgments (see also records regardless of sign) this increase is very apparent. Judgments are passed with absolute certainty on the basis of an average deviation of 5.2 per cent. in

tapping, but in color-naming 10.9 per cent. and in opposites 16.3 per cent. deviation is necessary to produce *A* judgments. The *C* judgments show this same increase without exception, and the *B* judgments differ only in the case of judgments of "worse" in opposites. The *D* judgments (pure guesses) show, as might be expected, no clear differences.

These differences in performance required for judgments of a given degree of confidence are not entirely a function of the variability of the trials in the three tests. The largest number of *A* judgments, as well as the smallest percentile variation for a given kind of judgment, comes in the tapping test, which is the least variable performance with all three individuals, in terms of per cent. variability. If the absolute variability be considered, the three tests all show practically the same mean variability, which varies from 1 to 2 seconds. Table VII. shows the average total time and the M.V. of 25 consecutive trials in each test, the trials being taken from the middle section of the experiment.

TABLE VII
SHOWING THE VARIABILITY OF THE TESTS

Test	<i>H</i>			<i>G</i>			<i>L</i>		
	Av.	M.V.	M.V.%	Av.	M.V.	M.V.%	Av.	M.V.	M.V.%
Tapping.....	40.3	1.5	3.7	40.1	2.0	5.0	36.5	1.0	2.7
Color-naming....	28.9	2.0	6.6	27.3	1.4	5.1	27.7	2.0	7.2
Opposites.....	30.3	2.0	6.6	26.6	1.8	6.7	23.7	1.6	6.7

This progression is doubtless partly dependent on decrease in the objectivity and automatic character of the three kinds of work. The more automatic and motor the work the greater the precision of the judgment of efficiency of performance. As the task comes to involve a greater proportion of more strictly mental work (association, memory, discrimination, choice, etc.) the judgments delivered with a given degree of confidence come to require larger and larger variations. Does this change involve a shift in the criteria (as for example, a shift from estimates of mere duration to reliance on affective processes, feelings of ease, smoothness, pleasantness, etc.)? Or does it involve merely a greater degree of some fairly constant criterion or criteria-complex? Is it perhaps due to the mere fact that there is better opportunity to observe the efficiency of an automatic process since it requires little attention itself? Systematic introspection during such an experiment would doubtless throw interesting light on the basis of the feeling of efficiency, and perhaps on the affective consciousness generally. Comparison of the judgments of a witness with judgments of the performer would be especially inter-

esting. The following two chapters will report an experiment in which these additional factors were studied.

4. In all three tests the various degrees of confidence have a very constant ratio of correctness. About 60 per cent. of the *D* judgments, 70 per cent. of the *C* judgments, 80 per cent. of the *B* judgments, and 98 per cent. of the *A* judgments, are correct. Fullerton and Cattell point out that these ratios of correctness do not measure the intensity or amount of the *feeling of confidence*. The truth of this statement is obvious when one reflects that 50 per cent. of the judgments should be correct by mere chance. Perhaps a fairer measure of the amount of confidence is secured by subtracting this 50 per cent. chance correctness from each total correctness, thus leaving the various degrees of confidence as represented by magnitudes *A* (48), *B* (30), *C* (20), *D* (10). This would make the zero point the amount of confidence possessed by a judge who had absolutely no knowledge of what had happened. A still fairer measure would perhaps be the P.E. required for the given per cent. correctness. That there is, in the present experiments at least, a greater distance between the feeling of absolute certainty (*A*) and the first degree of uncertainty (*B*) than there is between the various degrees of uncertainty, agrees with the writer's own introspections. This is also borne out by the fact that the variations underlying these degrees of confidence do not increase by equal steps, but almost by equal multiples. The average deviation of the *C* judgments, regardless of sign, is about twice that of the *D* judgments, that of the *B*'s twice that of the *C*'s, and that of the *A*'s twice that of the *B*'s, if the percentile deviations be considered. If the absolute deviations be taken, they increase by 50 per cent. increments from *D* to *C* and from *C* to *B*, but the step from *B* to *A* represents an addition of 100 per cent. over the *B* judgments.

By referring to tables for determining the P.E. from the percentage of right cases and amount of difference, as in the method of right and wrong cases, we get:

Degree of confidence.....	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
Average difference—per cent.....	10.8	5.2	3.2	1.9
Per cent. right judgments	98	81	73	59
Diff./P.E.....	3.05	1.30	.91	.34
P.E.....	3.1	4.0	3.5	5.6

That is to say, the average probable error, the amount of variation which will be judged correctly in 75 per cent. of the cases, is about 4 per cent. of the "usual" record.

5. Individual differences in the use of the various degrees of confidence, in the percentile correctness, and in the probable error, have

been pointed out by Fullerton and Cattell and by Henmon. The present study of but three observers does not afford sufficient material for individual comparisons of any reliability. The numbers of cases of a given sort vary from individual to individual and in some instances are small. With respect to the amount of variation on which the various judgments are based, the results are much the same for all observers in those cases in which the number of trials is large enough to make comparison reliable. The same thing must be said of the correctness of judgment. Such differences as are found are either small or are in no consistent direction. With respect to the distribution of judgments ("better" or "worse") in tapping no individual differences are present. The judgments of "better" are somewhat in excess, but so are the actual cases of superior performance, to a slight degree. In color-naming the judgments of *G* and *L* are skewed considerably toward the "worse" side, but the actual cases are distributed in much the same way with these two observers. With *H* the actual cases of each sort are equal and the distribution of judgments is uniform. In the case of opposites much the same situations are present.

Summary

1. The study of the conditions, validity, and laws of judgments of personal efficiency offers a fruitful field of inquiry, with respect to the psychology of judgment, the learning process, affective consciousness, the psychology of work, and individual differences.

2. In the tests examined, an individual's judgments of the efficiency or inefficiency of his own performance possess a degree of correctness which varies with his degree of confidence. In this respect judgments of performance resemble judgments of sensory discrimination and of recognition memory. The relative per cent. correctness of the four degrees of certainty are 98, 80, 70, and 60. Pure guesses are more likely to be right than wrong.

3. The feeling of efficiency arises on slighter provocation than does the feeling of inefficiency. Judgments of greater efficiency, having a given degree of confidence, are based on smaller variations in performance than are equally confident judgments of inferior performance.

4. Judgments of "better than usual" show nearly as high per cent. correctness as do judgments of "worse than usual," although the former are based on variations about one half as great as those on which the latter are based.

5. There is a slight predisposition toward the delivery of judgments of "better," the distribution being, however, on the average,

within 5 per cent. of the actual ratio of occurrence of superior and inferior trials.

6. Progressively larger variations in performance are required as the basis for judgments of a given degree of confidence as one passes from an automatic, objectively observable, motor performance (such as tapping), through work involving perceptual reactions (color-naming), to work of a more strictly mental and less objectively observable character (opposites).

7. No evidence is here afforded on the question as to whether this decreasing precision of judgment depends on a shift in criteria (as from estimates of duration to reliance on affective processes) or on the greater intensity or clearness of some fairly constant criterion or criteria-complex.

8. A variation of 4 per cent. from "usual" will be judged correctly in 75 per cent. of the cases in which it occurs.

9. Judgments of *A*, *B*, *C*, and *D* degrees of confidence show a per cent. correctness which is respectively 48 per cent., 30 per cent., 20 per cent., and 10 per cent. greater than would result from chance estimates. ($\Delta/P.E = 3.05, 1.30, .90, .34$). These ratios are confirmed by introspection as approximate measures of the intensity of the "feeling of certainty" in the four cases. These ratios do not differ essentially from the corresponding degrees of correctness of similar judgments of sensory discrimination and recognition memory. They depend, in part, however, on the character and difficulty of the task and on the range of variation in stimulus, stimulus difference, and time of performance.

10. The number of observers is insufficient for the determination of the nature or degree of individual differences.

CHAPTER II

PERCEPTUAL CRITERIA OF JUDGMENTS OF EFFICIENCY

IN daily life these judgments of personal efficiency are frequently expressed. A worker asserts that his work is "going unusually well," that he is "in fine form," or, on the other hand, that he is "not himself," that his work is not "up to its usual standard," etc. Not only does the performer himself pass such judgments, but witnesses may make similar remarks. These judgments may be delivered with varying degrees of confidence, ranging from pure guessing to absolute assurance. They are passed on muscular work involving only strength or endurance, on work requiring more or less coordination, on work involving sensory discrimination and perceptual reaction, and on more exclusively mental work. Shovelling coal, riding a bicycle, playing tennis, target shooting, mathematical calculation, and writing sonnets represent such gradations in daily life.

In many of these concrete situations the judgments of personal efficiency may be determined or supported by reference to the objective result of the work,—the wages earned, the score attained, etc. In such cases we should perhaps speak of "inferences" rather than of "judgments." But even in the absence of knowledge of the objective results a worker may estimate the efficiency of his work, and in these cases he does it by some direct process which seems, before analysis at any rate, to be correctly described as "judgment" of the most primary sort. Such judgments are often said to be the expression of "feelings,"—feelings of efficiency, of inefficiency, etc.

In the preceding chapter was reported a study of the distribution, confidence, and accuracy of such judgments. The present chapter presents the results of a further study designed to investigate the characteristics and criteria of these judgments, the way in which these features vary with the nature of the task, the effects of practise on the correctness of the judgments, and the relation, in all these respects, between judgments of one's own performance and judgments of the work of another person.

Four observers have taken part in the experiment, two men and two women, the two men being professional psychologists, one of the women an experienced psychological observer, the other a beginner. The work consisted in the repeated performance of four standard laboratory tests,¹ as follows:

¹ For further discussion of the nature, technique and significance of these

(a) *Color-naming*,—the Woodworth-Wells blank, containing 20 repetitions of each of 5 colors, the four positions of the card being used in succession.

(b) *Naming Opposites*,—a list of 50 adjectives, the antonyms to which were to be given as quickly as possible. The list was one used by the writer in previous studies, the average time of naming the opposites ranging from 2 to 5 seconds per word. The 50 words occurred in chance order, the order being changed at each trial.

(c) *Cancellation*,—crossing out the 3's and 5's from the Woodworth-Wells form of this test, the first 10 lines, containing 50 repetitions of each digit, being used.

(d) *Addition*,—adding 17 mentally to each of 50 two-place numbers and calling out the correct answer. The numbers occurred in changed random order at each trial.

The time of performance was taken, in fifth-seconds, for each trial, the quantity and quality of the work being maintained constant. Each observer made 104 trials at each test, the first 4 trials being considered preliminary. After the completion of the trial, and before the operator had recorded or even noticed the time measurement, both performer and operator judged the performance to have been either "better than usual" or "worse than usual," and assigned to the judgment one of four degrees of confidence,—*A*, *B*, *C*, or *D* (*A* representing absolute certainty, and *D* a mere guess). Both judgments were recorded independently, after which the objective measurement was recorded. This procedure thus yields, for each of the four tests, 100 judgments from each of four performers and 100 judgments from each of four witnesses,—a total of 3,200 judgments. The experiment occupied a two-hour session on each of 9 successive days, and 10 to 12 trials of each test were made at each sitting.

Toward the close of the experiment each subject was given the following schema for systematic introspection. The two arrangements of criteria were made on separate occasions, the first on the eighth and the second on the ninth day. After the completion of the experiment each observer was asked to answer the supplementary questions.

SCHEMA FOR INTROSPECTION

- A. Feelings of ease and comfort or of strain and uncertainty as the test proceeds.
- B. Feelings of pleasantness and satisfaction or of unpleasantness and dissatisfaction, either during the test or after its completion.
- C. The perception of the smoothness and regular flow or of the roughness and irregularity of the performance.

tests, and their usefulness as psychological instruments, the reader is referred to the writer's monograph, "The Influence of Caffein on Mental and Motor Efficiency," ARCHIVES OF PSYCHOLOGY, No. 22. 1912. Science Press.

- D. Direct estimate of the total time interval or duration of the test from beginning to end, regardless of what happens during the performance of the test.
- E. Perception of the speed or rate of succession of the separate acts which the test involves (as each word, each problem, etc.).
- F. Inference based on the number or amount of specific mistakes, hesitations, successes, observed during the test or remembered after its completion.
- G. Feelings of surprise, or of fulfilled or unfulfilled expectation, when the end of the test is reached.
- H. Unanalyzable and indefinable feeling of efficiency or of inefficiency.
- I. Any other specific criteria which you may have noted.

QUESTIONS ON THE SCHEMA

1. Think over the way in which you judge *your own performance* in each of the tests. Arrange the above factors in the order of their importance with respect to the degree to which they constitute the basis or criteria for your judgments of your own work. Place the most important first, then the next in importance, etc. Do this separately for each of the four tests.

2. Now think over the way in which you judge the *performance of another person*, and arrange the above criteria in the order of their importance, separately for each of the four tests, as was done in question 1.

SUPPLEMENTARY QUESTIONS

1. When do you feel the greater security or certainty,—when judging your own performance or when judging that of another person?

2. In which case do you think you can detect smaller changes or variations in efficiency of performance,—when judging yourself or when judging another person, in these tests?

3. In which of these four tests do you think your judgments are delivered with the greatest degree of confidence? Arrange the four tests in order of decreasing confidence, both for when judging yourself and for when judging another person.

4. In which of the tests do you believe you can detect the smallest changes in performance? Arrange the four tests in order, for this point, as in the preceding question.

5. When judging your own performance and that of another, which of the following is or are true?

(a) A judgment is made tentatively during the performance and this judgment is modified and corrected as the test proceeds, the judgment thus being ready at the moment when the test is completed.

(b) No judgment is made until the test is all completed, when the judgment is formed by thinking back over the test as a whole, as it was performed on the given occasion.

(c) At the end of the test the judgment simply comes, of its own accord, and fully formed. It is not made tentatively during the test, nor is it necessary to think back over the particular performance.

The present paper will present the results of this systematic introspection, an examination of the total per cent. correctness of the judgments, a statement of the influence of practise on correctness, a

TABLE VIII
INDIVIDUAL ARRANGEMENTS OF THE CRITERIA OF JUDGMENT

The Test	Order when Judging Self Observers					Order when Judging Another Person Observers					
	Position	H,	P,	L,	R	Position	H,	P,	L,	R	
Color naming:	1	C	E	E	F	1	C	C	F	H	
	1	E	C	F	A	2	E	E	C	C	
	3	A	F	C	C	3	F	F	D	A	
	4	F	A	B	D	4	G	D	E	E	
	6	G	B	A	H	5	D	{ A	G	F	
	6	B	H	D	B	6	A		{ B	H	D
	7	D	{ D	{ G	G	7	B			{ A	B
	8	H			{ G	{ H	E	8	H		{ B
Naming opposites:	1	E	E	F			H	1	E	C	
	2	A	C	C	C	2	C	E	C	C	
	3	C	F	G	E	3	D	F	D	E	
	4	F	A	B	B	4	F	D	E	F	
	5	B	B	E	G	5	G	{ A	G	H	
	6	G	H	A	F	6	A		{ B	{ H	B
	7	D	{ D	{ D	D	7	B				{ G
	8	H			{ G	{ H	A	8	H	{ H	
Cancellation:	1	E	C	A			C	1	E		C
	2	C	F	E	E	2	F	E	F	F	
	3	F	E	F	F	3	G	F	C	E	
	4	D	A	H	A	4	A	D	D	C	
	5	A	H	B	B	5	B	{ B	G	B	
	6	B	{ B	{ D	D	6	C		{ G	{ H	D
	7	G			{ C	{ D	G				7
	8	H	{ G	{ G			H	8	H	{ H	{ B
Adding:	1	A			F	E	C	1	E		
	2	E	E	A	H	2	G	E	E	A	
	3	C	C	B	G	3	F	F	C	C	
	4	F	A	C	E	4	C	D	D	B	
	5	D	B	F	F	5	A	{ A	G	E	
	6	G	H	G	B	6	B		{ B	{ H	F
	7	B	{ D	{ D	D	7	D				{ G
	8	H			{ G	{ H	A	8	H	{ H	

Brackets indicate criteria not used.

comparison of the process of judging one's self with that of judging another person, and some points on individual and test differences.

The Criteria of Judgment.—The eight items included in the schema proved to be a complete enumeration of the criteria used by all four observers. These eight criteria being arranged in order of importance by each observer, for each test, and both for judging as performer and for judging as witness, the final position of importance for each criterion is determined by averaging the four arrangements

for the given situation. The individual orders are given in Table VIII. The average positions of the eight criteria are given in Table IX.

It is clear at once, from Table IX., that criteria *E* (perception of speed or rate of succession of the separate elements), *C* (perception of the smoothness and regular flow or of the roughness and irregularity of performance), and *F* (inference based on number and amount of specific mistakes, hesitations, successes, etc.) are considered, and in the order here given, the most important criteria, both for personal judgments and for judgments as witness. This is further confirmed by observation of the number of times each criterion was reported "not used," out of a total of 32 possible situations (4 tests, 4 observers, as performer and as witness). The figures are as follows:

Criterion	Times Reported as Not Used
<i>A</i>	7
<i>B</i>	8
<i>C</i>	0
<i>D</i>	4
<i>E</i>	0
<i>F</i>	0
<i>G</i>	8
<i>H</i>	8

TABLE IX

FINAL AVERAGE POSITIONS OF ALL CRITERIA

When Judging One's Own Performance

Criterion	Colors	Opposites	Cancellation	Adding	Grand Av.
<i>A</i>	3.5	5.0	3.5	3.8	3.9
<i>B</i>	5.3	4.5	4.0	5.3	4.8
<i>C</i>	2.3	2.3	2.8	2.8	2.5
<i>D</i>	6.0	7.0	5.8	6.5	6.3
<i>E</i>	3.0	2.5	2.0	2.3	2.4
<i>F</i>	2.5	3.5	2.8	3.8	3.1
<i>G</i>	6.8	5.5	7.5	5.8	6.4
<i>H</i>	6.8	5.8	6.3	6.0	6.2

Final Order, *E-C-F-A-B-H-D-G*

When Judging the Performance of Another Person

<i>A</i>	5.3	4.8	6.5	4.8	5.3
<i>B</i>	7.0	6.8	5.8	6.0	6.4
<i>C</i>	1.5	1.8	3.5	2.8	2.4
<i>D</i>	4.5	4.5	5.3	4.5	4.7
<i>E</i>	2.5	2.5	1.8	2.5	2.3
<i>F</i>	3.0	3.0	2.3	3.3	2.9
<i>G</i>	6.0	6.0	5.3	5.5	5.7
<i>H</i>	5.8	6.8	5.8	5.8	6.3

Final Order, *E-C-F-D-A-G-H-B*

Criteria *C*, *E*, and *F* are the only ones never reported "not used." The direct estimate of total time interval or duration (*D*) is given a higher value (4.7) when judging another than when judging one's self (6.3). Feelings of surprise (*G*) show a similar difference, which is, however, only slight (5.7 and 6.4). Feelings of pleasantness or unpleasantness (*B*) have a much higher value when judging one's self (4.8) than when judging another person (6.4). Unanalyzable feelings of efficiency or of inefficiency (*H*) average only slightly higher when judging one's self and as a matter of fact only the untrained observer ever places this criterion higher than the sixth position.

In general, then, the affective processes do not, in the opinion of these four observers, play any considerable rôle as criteria of judgments of efficiency in these tests. The criteria chiefly relied on are directly perceptual in character (speed, smoothness or roughness) or are inferences from particular delays or successes. Trained observers do not report an "unanalyzable feeling of efficiency," but point to specific criteria of a perceptual character; nor is the estimate of total time interval or duration important. The great difference between the positions of *E* (speed) and *D* (duration) seems to indicate a probable direct and independent basis for judgments of *speed* of performance, as is also found to be the case with judgments of the characteristics of voluntary movements.² Because of the importance of these perceptual factors, the judgments of the performance of another person are based on the same criteria as are those of one's own work.

Correctness of the Judgments.—All four observers report greater confidence when judging themselves, and believe themselves to be more sensitive to changes in their own performance than in that of another person. Table X. shows the per cent. correctness of the judgments in all situations. In computing these results, the median of the five trials preceding a given test was used as the standard of comparison, or as a measure of "usual" performance. By *usual* is thus meant the median record of the *half-day's work* immediately preceding the trial in question. This standard was adopted after questioning the observers as to the meaning which the term "usual" had for them, and its use accords with the introspections of all four observers. In the table the degree of confidence of the judgments is ignored, since this matter will be taken up in the following chapter. The judgment is counted correct or incorrect according as the record did or did not differ, in the direction asserted, from the median of the

² See Hollingworth, "The Inaccuracy of Movement," pp. 40-62.

five preceding trials, regardless of both the amount of this deviation and the degree of confidence. In connection with this table three points are to be especially noted.

TABLE X
SHOWING THE PER CENT. OF CORRECT JUDGMENTS

Test	Observers				
	<i>H</i>	<i>P</i>	<i>L</i>	<i>R</i>	Average
Color-naming	72	67	65	67	68
Opposites	80	68	69	73	72
Cancellation	74	67	71	72	71
Addition	69	70	74	69	70
Averages	74	68	70	70	70

<i>When Judging Another Person</i>					
Color-naming	60	69	60	52	60
Opposites	67	62	66	64	60
Cancellation	59	79	67	61	66
Addition	73	67	80	63	71
Averages	65	69	68	60	65

1. Within any given judgment situation there are no considerable individual differences in correctness. Such differences as are present are not consistently individual.

2. Correctness when judging one's self is, on the average, only about 5 per cent. higher than when judging another person. This difference, such as it is, confirms the introspective reports of the four observers. Its slight amount bears additional witness to the perceptual character of the criteria of the judgments. Factors *E*, *C*, and *F* are as directly observable in estimating another's work as when judging one's own performance. The slight difference found may be accounted for in part by the greater degree of attention given to the process when one judges himself.

3. This average difference of about 5 per cent. is due to the first three tests on the list (color-naming, opposites, and cancellation). The per cent. superiority in the correctness of the personal judgments in the various tests is +12 per cent. for opposites, +8 per cent. for color-naming, +5 per cent. for cancellation, and -1 per cent. for adding. For the individual subjects these differences are as shown in Table XI. If these small differences are at all significant, they probably indicate only differences in the degree to which one is able to take an objective attitude toward his own performance, and color-naming and opposites would thus seem to involve processes

more reflex in character than those involved in cancellation and adding.

TABLE XI

SHOWING FOR EACH SUBJECT AND EACH TEST THE SUPERIORITY OF THE CORRECTNESS OF PERSONAL JUDGMENTS OVER THAT OF JUDGMENTS OF THE PERFORMANCE OF ANOTHER PERSON, IN PER CENT.

Observer	H	P	L	R
Color-naming.....	13	6	3	9
Opposites.....	12	- 2	5	15
Cancellation.....	15	-12	4	11
Addition.....	-4	3	-6	6

Practise Effects.—Table XII. gives the per cent. of correct judgments for each section of 20 trials. There is no considerable practise gain in correctness in the separate tests nor with the different observers. The fourth section (trials 61 to 80) tends to show greatest correctness, and quite uniformly. But in the personal judgments there

TABLE XII

SHOWING THE EFFECT OF PRACTISE ON CORRECTNESS OF JUDGMENT. THE FIGURES INDICATE THE TOTAL NUMBER OF CORRECT JUDGMENTS DELIVERED BY ALL FOUR OBSERVERS, IN EACH SITUATION

Trials	Color Naming			Opposites			Cancellation			Addition		
	Perf.	Wit.	Tot.	Perf.	Wit.	Tot.	Perf.	Wit.	Tot.	Perf.	Wit.	Tot.
1-20	48	41	89	59	51	110	53	54	97	56	52	108
21-40	53	49	102	54	45	99	53	53	106	52	50	102
41-60	45	44	89	55	50	105	52	52	104	61	58	119
61-80	59	54	113	57	54	111	64	58	122	54	57	111
81-100	56	48	104	53	49	102	53	54	107	57	60	117

is, aside from this, no gain. In judgments as witness there is, if the grand totals be considered, a fairly well marked increase in correctness in the successive sections of the experiment. Further than stating these points it is difficult to analyze out the practise factor. The real gain is probably in all cases greater than the figures reveal, because, as the experiment proceeded, the magnitude of the variations from trial to trial grew smaller and smaller, as the result of practise in the tests themselves. Meanwhile the "usual" record also became better and better. The same per cent. correctness (and, as witness, a higher correctness) is maintained in spite of this decrease in absolute variability. On the other hand, this is what would be expected if something like Weber's law holds in such judgments.

The slightly superior correctness of the personal judgments is present in all five sections of the experiment (see Tables XII. and XIII.), but it decreases somewhat as the later sections are passed through. This decrease seems to depend solely on such practise gain

as comes when the judgments are directed toward the work of another person.

TABLE XIII

GRAND TOTAL CORRECTNESS, ALL TESTS, ALL OBSERVERS

Trials	Judging Self	Judging Another	Totals
1-20	216	188	404
21-40	212	197	409
41-60	213	204	417
61-80	234	223	457
1-100	219	211	430

Showing Effect of Practise on the Correctness of the Judgments. Witness gains, approximating finally the correctness of the performer.

Formulation of the Judgments.—Individuals differ somewhat in their methods of formulating the judgments, and the process varies also with the test and with the judgment situation. Thus observer *H* reports: "When judging myself, no judgment is usually formed until the test is completed, in which case the judgment may either seem to come of its own accord, fully formed, or it may require thinking back over the trial and comparing it with other trials. But when judging another person a tentative judgment is usually made early in the performance and this judgment is modified as the test proceeds, and is ready for delivery at the moment the test is completed. This is particularly true of cancellation and of addition. In color-naming and opposites it is less true."

Similarly, observer *L* reports: "I seem to form judgments in all three ways suggested, sometimes in one way, sometimes in another." The other two observers describe themselves as having relied chiefly on the method of tentative formulation and modification, regardless of the test or of the judgment situation (as performer or as witness).

Summary

The chief results of the study may be summarized as follows:

1. The important criteria of judgments of efficiency in these tests are either directly perceptual in character or are inferences from such data. Affective processes do not play an important rôle.
2. A direct and independent basis or set of sensory criteria for judgments of *speed* of performance is indicated.
3. The same criteria are relied on when judging one's own efficiency as when judging that of another person.
4. Direct estimate of duration and feelings of surprise are more important when judging another than when judging one's self. With feelings of pleasantness and unpleasantness and with unanalyzable feelings of efficiency or inefficiency the reverse is the case.

5. Trained observers do not report unanalyzable feelings, but point to specific criteria of a perceptual character.

6. Judgments of one's own work tend to be only slightly better, from the point of view of correctness, than judgments of the work of another person. This superior correctness of the personal judgments varies somewhat with the test. It is greater for color-naming and opposites than for cancellation and addition.

7. Practise results in an absolute increase in correctness in the case of judgments as witness. Personal judgments show no absolute gain but the initial per cent. correctness is maintained along with a decrease in the absolute variability of the trials. There is thus in both cases a real improvement, which is greater than the figures indicate.

8. The process of judgment formulation, as introspectively described, differs with the individual, with the test, and with the judgment situation.

CHAPTER III

PERFORMER AND WITNESS AS JUDGES OF EFFICIENCY

THE two previous chapters have presented results bearing on the judgment of personal efficiency in a work process, the characteristics, reliability and laws, and the basis or criteria of these judgments. In the first chapter it was shown (1) that an individual's judgment of his own efficiency in a task just completed possesses a degree of correctness which varies in a definite and measurable way with his feeling of confidence in the judgment; (2) that judgments of "better than usual" are nearly as often correct as are judgments of "worse than usual," although the former do tend to be somewhat in excess of the number of actual cases; (3) that the magnitude of the average constant variation required as the basis of judgments of a given degree of confidence varies with the nature of the task; and (4) that judgments of "better" arise on slighter provocation than do judgments of "worse."

The second chapter gave the results of an introspective study of the judgment of efficiency, both when judging one's self and when judging the performance of another person. It was here indicated that (1) the important criteria relied on in making these judgments are either directly perceptual in character or inferences from such data; (2) that affective processes do not play an important rôle as criteria of these judgments, and that unanalyzable feelings of efficiency or feelings of inefficiency are not reported; (3) that the same criteria are relied on when judging one's own performance as when judging that of another person; (4) that the specific criteria and the process of formulating the judgment vary with the task and with the judgment situation, and (5) that one's judgments of his own performance are only slightly more correct than his judgments of the work of another person, the latter judgments improving somewhat in correctness as the result of practise.

The present chapter reports a continuation of this series of investigations, designed to check up the previous results by securing a larger number of judgments from more observers and in new tasks, and to make a thorough quantitative and qualitative comparison of the judgments of performer and of witness. Since the method used here was identical with that described in the earlier studies no detailed account of it need be given here. Four observers, two men

and two women, took part in the experiments. Four tests were employed, described in earlier papers: *Color-naming*, *Naming Opposites*, *Cancellation*, and *Addition*. The data discussed in this chapter were secured in connection with the experiment described in Chapter II.

The time of performance was taken in fifth-seconds. After four preliminary trials each observer made 100 further trials. After each trial, and before the operator had noted the record, both performer and operator judged the performance to have been either "better than usual" or "worse than usual," and each assigned to his or her judgment one of the four degrees of confidence (*A*, *B*, *C*, or *D*). Both judgments were independently recorded, and after this was done the objective measurement was noted by the operator only. Each person served in turn as operator and as performer. This procedure gives 100 judgments from each of four performers and 100 from each of four witnesses, the two sets of judgments referring to the same records. Since there were four tasks this gives a total of 3,200 judgments. The experiments occupied a two-hour period on

TABLE XIV

SHOWING THE AMOUNT OF PRACTISE GAIN IN THE VARIOUS TESTS

Test	Observer	Average of First 10 Trials (Sec.)	Average of Last 10 Trials (Sec.)	General Average (Sec.)	Gain (Sec.)
Color-naming:	<i>R</i>	34	36	35	-2
	<i>L</i>	51	45	48	6
	<i>P</i>	45	40	43	5
	<i>H</i>	42	38	40	4
Opposites:	<i>R</i>	28	23	26	5
	<i>L</i>	25	22	24	3
	<i>P</i>	50	34	42	16
	<i>H</i>	32	28	30	4
Cancellation:	<i>R</i>	75	55	65	20
	<i>L</i>	56	46	51	10
	<i>P</i>	60	40	50	20
	<i>H</i>	54	40	47	14
Addition:	<i>R</i>	90	52	71	38
	<i>L</i>	86	50	68	36
	<i>P</i>	100	58	79	42
	<i>H</i>	83	60	72	23

each of 9 successive days, 10 to 12 trials of each task being made at each sitting by each person.

In computing results the median of the five trials preceding the given record was used as the standard of comparison, or as the meas-

ure of "usual" performance. This standard was adopted after questioning the observers as to the meaning which the term "usual" had for them. By "usual" is thus meant the median record of the half-day's work immediately preceding the trial in question. It may be well to point out that this method was used (rather than, for instance, comparison with the preceding trial) in order to make the experiment as nearly as possible comparable with daily life, in which our impressions and verdicts of momentary efficiency of ourselves or of others are usually expressed in these general terms.

TABLE XV

ABSOLUTE DEVIATIONS FROM USUAL. JUDGING SELF. GIVING ALSO THE
RELIABILITY

Test	Obs.	Better				Worse			
		-A	-B	-C	-D	+A	+B	+C	+D
Colors:	H A.S.D. 1	-2.5	-1.5	-1.0	-2.2	4.9	3.0	1.2	-0.4
	P.E.	.4	.7	.5	.7	.2	.6	.8	.2
	P A.S.D.	-1.7	-0.8	-0.8	0.7	5.2	4.2	0.6	—
	P.E.	.5	.4	.3	.6	.8	.6	.6	—
	R A.S.D.	-0.5	-2.2	-1.1	-0.9	6.8	0.8	1.9	5.3
	P.E.	.8	.7	.4	.9	1.0	.3	.8	1.0
	L A.S.D.	-4.6	-3.2	-0.6	-1.2	6.0	2.7	0.9	0.3
	P.E.	.6	.8	.6	.9	1.3	1.4	.7	.6
Opps.:	H A.S.D.	-4.0	-3.5	-2.9	-0.1	4.5	3.4	0.8	-0.4
	P.E.	.4	.5	.4	.7	.6	.6	.4	1.1
	P A.S.D.	-3.2	-1.4	0.3	0.1	7.4	3.6	0.5	—
	P.E.	.6	.2	.3	.6	1.0	.7	.7	—
	R A.S.D.	-1.6	-0.7	0.1	0	5.4	1.6	0.9	2.3
	P.E.	.2	.2	.2	.2	.4	.3	.5	.7
	L A.S.D.	-2.4	-0.7	-0.1	1.1	1.9	3.4	3.2	0.7
	P.E.	.3	.3	.3	.4	.4	.5	.6	.3
Canc.:	H A.S.D.	-6.0	-3.6	-1.5	-1.5	4.9	5.8	0.1	0.9
	P.E.	.6	1.0	.7	.5	1.0	.7	.8	.5
	P A.S.D.	-4.6	-1.0	2.2	2.2	14.3	6.7	3.0	-0.8
	P.E.	.4	.4	.5	2.6	8.0	.8	.7	—
	R A.S.D.	-8.0	-4.4	-3.0	-1.5	8.6	.6	1.5	-3.3
	P.E.	—	.8	.9	.8	1.1	1.2	.4	1.4
	L A.S.D.	-6.4	-1.9	0.8	-0.3	7.0	5.9	3.3	0.7
	P.E.	.8	.7	.7	.6	1.4	.8	.8	.8
Add.:	H A.S.D.	-9.8	-5.6	-5.7	-5.5	8.3	3.4	-1.2	-2.4
	P.E.	1.0	2.1	1.5	1.0	1.0	1.4	1.1	1.1
	P A.S.D.	-7.2	-1.3	-1.0	-0.8	5.9	2.0	0.8	1.7
	P.E.	.6	.7	.8	—	1.0	.8	.7	2.5
	R A.S.D.	-9.4	-2.8	-3.1	-3.2	3.6	0.2	0.4	-1.4
	P.E.	1.3	.9	.9	.3	1.4	1.1	.7	1.5
	L A.S.D.	-4.6	-1.9	-0.6	-1.7	7.7	7.7	2.9	1.9
	P.E.	.4	.4	.7	.9	1.8	3.3	1.4	.5

¹ A.S.D. = Average Stimulus Difference.

Three of the observers (*L*, *R*, and *H*) had had prolonged previous practise in *color-naming*. Observer *P* had not, but since repetition brings little improvement in this test the gains by the end of the experiment were very slight in all cases. The same three observers were practised in *opposites* but *P*, who was not, shows a gain of some 16 seconds by the end of the experiment. In the cases of *cancellation* and *addition* the amounts of previous practise were unequal. In all cases the four preliminary trials served to overcome the initial diffi-

TABLE XVI

ABSOLUTE DEVIATIONS FROM USUAL. JUDGING AS WITNESS. GIVING ALSO THE RELIABILITY OF THE MEASURES

Test	Obs.	Better				Worse			
		-A	-B	-C	-D	+A	+B	+C	+D
Colors:	<i>H</i> A.S.D. 2	-7.8	-4.8	-2.1	-0.3	9.8	1.7	0.3	0.8
	P.E.	.5	.9	1.0	1.1	.7	.5	.5	.4
	<i>P</i> A.S.D.	-3.8	-0.7	0.8	0.8	3.3	3.7	1.9	-0.3
	P.E.	.4	.5	.5	.7	1.8	1.3	.5	.4
	<i>R</i> A.S.D.	-1.1	-0.1	2.9	-4.2	3.4	-0.2	0.2	-1.4
	P.E.	.3	.5	.8	.7	1.2	.8	.7	—
	<i>L</i> A.S.D.	-1.3	-2.3	0.3	1.4	2.4	4.5	5.1	1.6
	P.E.	.6	.5	.5	.7	0	.3	.6	.9
Opps.:	<i>H</i> A.S.D.	-2.5	-0.9	-0.4	0	—	2.1	2.1	2.2
	P.E.	.6	.3	.3	.5	—	.8	.5	.4
	<i>P</i> A.S.D.	-0.5	-0.1	0.6	2.9	5.0	1.2	2.9	—
	P.E.	.3	.3	.5	1.1	0	.9	.8	—
	<i>R</i> A.S.D.	-0.8	-0.9	0.4	-1.8	4.8	5.0	2.4	2.8
	P.E.	.3	.4	.6	0	.5	2.1	.9	—
	<i>L</i> A.S.D.	-2.1	-1.3	-2.1	1.4	5.0	5.5	3.2	0.6
	P.E.	.5	.5	.7	.8	1.3	.5	.9	.7
Canc.	<i>H</i> A.S.D.	-7.5	-1.3	-3.0	0.5	7.8	1.6	2.4	-0.3
	P.E.	1.2	1.1	1.2	.5	.4	1.0	.7	.6
	<i>P</i> A.S.D.	—	-4.4	-2.5	-1.9	11.3	5.4	1.8	3.5
	P.E.	—	.5	.6	1.3	2.1	1.1	.8	1.4
	<i>R</i> A.S.D.	-3.8	-0.7	0.7	2.8	—	-0.6	0.5	0.8
	P.E.	1.1	.5	.9	.6	—	1.1	1.6	—
	<i>L</i> A.S.D.	-4.1	0.1	-0.8	-0.3	11.6	-1.1	3.8	-0.4
	P.E.	.6	.7	.6	.6	1.6	3.6	.7	.8
Add.:	<i>H</i> A.S.D.	-5.9	-4.2	-2.7	-4.0	6.3	2.0	0	-0.9
	P.E.	.2	.7	1.3	.9	2.4	.9	.6	1.2
	<i>P</i> A.S.D.	-9.7	-4.0	0.1	-4.7	12.9	-0.8	-0.6	—
	P.E.	1.9	.6	.5	.5	4.0	1.4	.9	—
	<i>R</i> A.S.D.	-4.0	-2.5	0	-1.6	7.9	-0.7	1.3	-0.3
	P.E.	.7	.8	.9	—	1.0	1.0	.9	2.9
	<i>L</i> A.S.D.	-6.1	-7.5	-5.0	-3.6	11.5	2.9	3.8	1.0
	P.E.	1.1	1.4	1.1	.7	1.2	1.7	1.4	1.2

² A.S.D. = Average Stimulus Difference.

TABLE XVII

JUDGING SELF

Average Constant Deviations from Usual, in Terms of Per Cent. of Average

Test	Average Record	Obs.	-A	Better -B	-C	-D	+A	Worse +B	+C	+D
Colors:	43	H	- 5.8	- 3.5	-2.3	-5.1	11.4	7.0	2.8	-0.9
	40	P	- 5.8	- 2.0	-2.0	1.7	13.0	10.5	1.5	—
	35	R	- 1.4	- 6.3	-3.1	-2.6	19.4	2.3	5.4	1.4
	48	L	- 9.6	- 6.6	-1.2	-2.5	12.5	5.6	1.9	0.6
	Averages.....		- 5.2	- 4.6	-2.2	-2.1	14.3	6.4	2.9	0.4
Total No. of cases.....			44	76	71	50	28	52	50	29
Opposites:	30	H	-13.6	-11.7	-9.7	-0.3	15.0	11.3	2.7	-1.3
	42	P	- 7.6	- 3.3	0.7	0.2	17.6	8.6	1.2	—
	26	R	- 6.4	- 2.8	0.4	0	21.6	6.4	3.6	9.2
	24	L	- 9.6	- 2.8	-0.4	4.4	7.6	13.6	12.8	2.8
	Averages.....		- 9.3	- 5.1	-2.2	1.1	15.5	10.0	5.1	3.6
Total cases.....			75	91	62	29	33	41	43	26
Cancellation:	47	H	-12.7	- 7.6	-3.2	-3.2	10.5	12.3	.2	1.9
	50	P	- 9.2	- 2.0	4.4	4.4	28.6	13.4	6.0	-1.6
	65	R	-12.3	- 6.7	-4.6	-2.3	13.2	0.9	2.3	-5.1
	51	L	-12.8	- 3.8	1.6	-0.6	14.0	11.8	6.6	1.4
	Averages.....		-11.7	- 5.0	-0.4	-0.4	16.6	9.6	3.8	0.8
Total cases.....			52	82	68	44	30	33	56	35
Adding:	72	H	-13.6	- 7.8	-7.9	-7.6	11.5	4.6	-1.7	-3.3
	79	P	- 9.1	- 1.6	-1.3	-1.0	7.5	2.5	1.0	-2.1
	71	R	-13.2	- 3.9	-4.3	-4.5	5.0	0.3	0.6	-1.9
	68	L	- 6.8	- 2.8	-0.9	-2.5	11.3	11.3	4.3	2.8
	Averages.....		-10.7	- 4.0	-3.6	-3.9	8.8	4.7	1.0	-1.1
Total cases.....			73	86	48	25	39	52	43	34

culties and to bring the performer close to the secondary slope of the practise curve. Table XIV. gives the averages of the first 10 trials (excluding the preliminaries) and of the last 10 trials, the average of these two averages, and the difference between them, thus affording an approximate statement of the general tendency to gain for each individual.

In the case of each trial the difference between the record made and the appropriate measure of "usual" was found. These differences were then assembled according to the judgments passed on them, the judgments of "better" and of "worse," each with the four degrees of confidence, being tabulated separately. The average constant deviation from usual was then computed for each type of judgment, for each test, and for each individual, both as performer and as witness. Tables XV. and XVI. give these absolute constant deviations, along with their variability.

In these tables, as in those which follow, the sign (—) means “better” (*i. e.*, requiring less time than usual) and the sign (+) means “worse” than usual.

In Tables XVII. and XVIII. these absolute deviations have been transformed into per cent. of the average time of performance in the case of each person. This makes it possible to treat all the deviations as comparable magnitudes. In these two tables the deviations are assembled according to the test, and test-averages are also computed. In Tables XIX. and XX. the same measures are reassembled according to the individual, and individual averages are computed. Table XXI. represents the individual averages and the combined averages, for all types of judgment. Table XXII. presents the combined test averages for all types of judgment, and is a convenient summary of many of the most interesting results of the experiment. Table XXIII.

TABLE XVIII

JUDGING AS WITNESS

Average Constant Deviations from Usual, in Terms of Per Cent. of Average of the Time of the Performer

Test	Average		Better				Worse			
	Record	Obs.	-A	-B	-C	-D	+A	+B	+C	+D
Colors:	48	H	-15.6	-9.6	-4.2	-0.6	19.8	3.4	0.6	1.6
	35	P	-11.4	-2.1	2.4	2.4	9.9	11.1	5.7	-0.9
	40	R	-2.7	-0.2	7.2	-10.5	8.5	-0.5	0.5	-3.5
	43	L	-3.0	-5.4	0.7	3.3	5.6	10.5	11.9	3.7
Averages.....			-8.2	-4.3	1.5	-1.4	10.9	6.1	4.7	-0.2
Total cases.....			80	85	56	35	14	43	47	40
Opposites:	24	H	-10.0	-3.6	-1.6	0	—	8.4	8.4	8.8
	26	P	-2.0	-0.4	2.4	11.6	20.0	4.8	11.6	—
	42	R	-1.9	-2.1	0.6	-4.3	11.5	11.9	5.7	6.6
	30	L	-7.0	-4.3	-7.0	4.3	16.6	18.3	10.7	2.0
Averages.....			-5.2	-2.6	-1.4	2.9	16.0	10.8	9.1	5.8
Total cases.....			129	93	52	29	18	17	29	33
Cancellation:	51	H	-15.0	-2.6	-6.0	1.0	15.6	3.2	4.8	-0.6
	65	P	—	-6.8	-3.8	-2.9	17.4	8.3	2.8	5.4
	50	R	-7.6	-1.4	1.4	5.6	—	-1.2	1.0	1.6
	47	L	-8.7	0.2	-1.7	-6.4	24.7	-2.3	8.1	-0.9
Averages.....			-10.4	-2.6	-2.5	-0.7	19.2	2.0	4.2	1.4
Total cases.....			36	115	67	46	11	38	51	36
Adding:	68	H	-8.7	-6.2	-3.9	-5.9	9.2	2.9	0	-1.3
	71	P	-13.7	-5.6	0.1	-6.6	18.2	-1.1	-0.8	—
	79	R	-5.0	-3.2	0	-2.0	10.0	-0.9	1.6	-0.4
	72	L	-8.5	-10.5	-6.9	-5.0	16.0	4.0	5.3	1.4
Averages.....			-8.9	-6.4	-2.7	-4.9	13.3	1.2	1.5	-0.1
Total cases.....			62	86	58	23	28	47	65	31

TABLE XIX

INDIVIDUAL RECORDS, JUDGING SELF

In Terms of the Per Cent. Constant Deviation from Usual

Obs.	Test	Better				Worse			
		-A	-B	-C	-D	+A	+B	+C	+D
H:	Col.....	- 5.8	- 3.5	-2.3	-5.1	11.4	7.0	2.8	-0.9
	Opps.....	-13.6	-11.7	-9.7	-0.3	15.0	11.3	2.7	-1.3
	Canc.....	-12.7	- 7.6	-3.2	-3.2	10.5	12.3	0.2	1.9
	Add.....	-13.6	- 7.8	-7.9	-7.6	11.5	4.6	-1.7	-3.3
	Average.....	-11.4	- 7.6	-5.8	-4.0	12.1	8.8	1.0	-0.9
P:	Col.....	- 4.2	- 2.0	-2.0	1.7	13.0	10.5	1.5	—
	Opps.....	- 7.6	- 3.3	0.7	0.2	17.6	8.6	1.2	—
	Canc.....	- 9.2	- 2.0	4.4	4.4	28.6	13.4	6.0	-1.6
	Add.....	- 9.1	- 1.6	-1.3	-1.0	7.5	2.5	1.0	-2.1
	Average.....	- 7.5	- 2.2	0.4	1.3	14.2	8.7	2.4	-1.9
R:	Col.....	- 1.4	- 6.3	-3.1	-2.6	19.4	2.3	5.4	1.4
	Opps.....	- 6.4	- 2.8	0.4	0	21.6	6.4	3.6	9.2
	Canc.....	-12.3	- 6.7	-4.6	-2.3	13.2	0.9	2.3	-5.1
	Add.....	-13.2	- 3.9	-4.3	-4.5	5.0	0.3	0.6	-1.9
	Average.....	- 8.3	- 4.9	-2.9	-2.3	14.8	2.5	3.0	0.9
L:	Col.....	- 9.6	- 6.6	-1.2	-2.5	12.5	5.6	1.9	0.6
	Opps.....	- 9.6	- 2.8	-0.4	4.4	7.6	13.6	12.8	2.8
	Canc.....	-12.8	- 3.8	1.6	-0.6	14.0	11.8	6.6	1.4
	Add.....	- 6.8	- 2.8	-0.9	-2.5	11.3	11.3	4.3	2.8
	Average.....	- 9.7	- 4.0	-0.2	-0.3	11.3	10.6	6.4	1.9

gives the test averages for *A*, *B*, *C*, and *D* judgments regardless of sign, secured by averaging the thresholds for "better" and "worse" judgments for each degree of confidence. Table XXIV. shows the distribution of the judgments for all types of situation and indicates the per cent. correctness in each case. The remaining tables are described later. In the discussion which follows these tables will be referred to by number.

Results

1. *Judgments of "better" are based on smaller constant deviations in efficiency than are judgments of "worse."* Considering the average percentile results from the four tests combined (Tables XVII., XVIII., and XXII.) this is true (1) for all four observers, (2) for all four degrees of confidence, and (3) both when judging self and when judging the performance of another. The difference is somewhat greater when judging another than when judging one's own performance. The average amounts of change required as the basis for judgments of any given degree of confidence are almost twice as large when judging inefficiency as when judging efficiency.

TABLE XX
INDIVIDUAL RECORDS. JUDGING AS WITNESS

Obs.	Test	Better				Worse			
		-A	-B	-C	-D	+A	+B	+C	+D
H:	Col.....	-15.6	-9.6	-4.2	-0.6	19.8	3.4	0.6	1.6
	Opps.....	-10.0	-3.6	-1.6	0	—	8.4	8.4	8.8
	Canc.....	-15.0	-2.6	-6.0	1.0	15.6	3.2	4.8	-0.6
	Add.....	-8.7	-6.2	-3.9	-5.9	9.2	2.9	0	-1.3
	Average.	-12.3	-5.5	-3.9	-1.4	14.9	4.5	3.5	2.1
P:	Col.....	-11.4	-2.1	2.4	2.4	9.9	11.1	5.7	-0.9
	Opps.....	-2.0	-0.4	2.4	11.6	20.0	4.8	11.6	—
	Canc.....	—	-6.8	-3.8	-2.9	17.4	3.2	4.8	-0.6
	Add.....	-13.7	-5.6	0.1	-6.6	18.2	-1.1	-0.8	—
	Average.	-9.0	-3.7	0.3	1.1	16.4	2.2	5.3	-0.8
R:	Col.....	-2.7	-0.2	7.2	-10.5	8.5	-0.5	0.5	-3.5
	Opps.....	-1.9	-2.1	0.6	-4.3	11.5	11.9	5.7	6.6
	Canc.....	-7.6	-1.4	1.4	5.6	—	-1.2	1.0	1.6
	Add.....	-5.0	-3.2	0	-2.0	10.0	-0.9	1.6	-0.4
	Average.	-4.3	-1.7	2.3	-2.8	10.0	2.3	2.2	1.1
L:	Col.....	-3.0	-5.4	0.7	3.3	5.6	10.5	11.9	3.7
	Opps.....	-7.0	-4.3	-7.0	4.3	16.6	18.3	10.7	2.0
	Canc.....	-8.7	0.2	-1.7	-6.4	24.7	-2.3	8.1	-0.9
	Add.....	-8.5	-10.5	-6.9	-5.0	16.0	4.0	5.3	1.4
	Average.	-6.8	-5.0	-3.7	-0.9	15.7	7.6	9.0	1.6

TABLE XXI
COMBINED AVERAGES OF ALL TESTS

Obs.	Situation	Better				Worse			
		-A	-B	-C	-D	+A	+B	+C	+D
H:	Self.....	-11.4	-7.6	-5.8	-4.0	12.1	8.8	1.0	-0.9
	Witness....	-12.3	-5.5	-3.9	-1.4	14.9	4.5	3.5	2.1
P:	Self.....	-7.5	-2.2	0.4	1.3	14.2	8.7	2.4	-1.9
	Witness....	-9.0	-3.7	0.3	1.1	16.4	2.2	5.3	-0.8
R:	Self.....	-8.3	-4.9	-2.9	-2.3	14.8	2.5	3.0	0.9
	Witness....	-4.3	-1.7	2.3	-2.8	10.0	2.3	2.2	1.1
L:	Self.....	-9.7	-4.0	-0.2	-0.3	11.3	10.6	6.4	1.9
	Witness....	-6.8	-5.0	-3.7	-0.9	15.7	7.6	9.0	1.6
Average self.....		-9.2	-4.7	-2.1	-1.3	10.6	7.6	3.2	0
No. of cases.....		244	335	249	148	130	178	192	124
Average, witness.		-8.1	-3.9	-1.3	-1.0	14.5	5.0	4.9	1.7
No. of cases.....		307	379	233	133	71	145	192	140

When the four individuals are averaged for each test, as in Tables XIX., XX., and XXII., this law holds for all tests with the exception of *addition*. Here it holds only for *B* judgments of one's own performance and for *A* judgments as witness.

These results quite confirm the similar finding reported in the earlier experiment (see Chapter I.). It was there questioned whether this law results from a predisposition toward judgments of "better," since these judgments show a somewhat lower per cent. correctness than do judgments of "worse." The result does not follow from the possibility of larger variations in the direction of inferiority, since these variations are, as a matter of fact, no more frequent, and even if they were would affect only the *A* judgments, whereas the law holds for all degrees of confidence. The only other explanation suggested was that the criteria of judgments of "better" are either different, more numerous, or more definite and more clearly detected, and that for this reason the "feeling of efficiency" arises on slighter provocation (smaller changes in performance) than does the "feeling of inefficiency."

TABLE XXII
COMPARISON OF WITNESS AND PERFORMER

Test	Situation	Better				Worse			
		-A	-B	-C	-D	+A	+B	+C	+D
Col.:	Self.....	- 5.2	-4.6	-2.2	-2.1	14.3	6.4	2.9	0.4
	Witness.....	- 8.2	-4.3	1.5	-1.4	10.9	6.1	4.7	-0.2
Opps.:	Self.....	- 9.3	-5.1	-2.2	1.1	15.5	10.0	5.1	3.6
	Witness.....	- 5.2	-2.6	-1.4	2.9	16.0	10.8	9.1	5.8
Canc.:	Self.....	-11.7	-5.0	-0.4	-0.4	16.6	9.6	3.8	-0.8
	Witness.....	-10.4	-2.6	-2.5	-0.7	19.2	2.0	4.2	1.4
Add.:	Self.....	-10.7	-4.0	-3.6	-3.9	8.8	4.7	1.0	-1.1
	Witness.....	- 8.9	-6.4	-2.7	-4.9	13.3	1.2	1.5	-0.1
Grand average.....		- 8.6	-4.3	-1.7	-1.2	14.1	6.3	4.0	0.8
Total cases.....		551	714	482	281	201	323	384	264

For averages, for self and for witness, see end of Table XXI.

Some information on this point is offered by the introspective accounts of the relative importance of various criteria relied on in making these judgments. (See Ch. II.) Each observer was given, toward the close of the experiments, the list of criteria, and was asked at the end of the investigation to arrange these various criteria in order of importance, according to the degree to which the criteria were used in judging "better" and also in judging "worse."

POSSIBLE CRITERIA OF JUDGMENT

- A. Feelings of ease and comfort or of strain and uncertainty as the test proceeds.
- B. Feelings of pleasantness and satisfaction or of unpleasantness and dissatisfaction, either during the test or after its completion.
- C. Perception of the smoothness and regular flow or of the roughness and irregularity of the performance.
- D. Direct estimate of the total time interval or duration of the test from beginning to end, regardless of what happens during the performance of the test.

- E.* Perception of the speed or rate of succession of the separate acts which the test involves (as each word, problem, etc.).
- F.* Inference, based on the number or amount of specific mistakes, hesitations, successes, observed during the test or remembered after its completion.
- G.* Feelings of surprise, or of fulfilled or unfulfilled expectation, when the end of the test is reached.
- H.* Unanalyzable and indefinable feeling of efficiency or of inefficiency.
- I.* Any other specific criteria which you may have noted.

The following table shows the arrangements by each individual:

Criterion	Criteria of Better Observers					Criteria of Worse				
	<i>H</i>	<i>P</i>	<i>L</i>	<i>R</i>	Av. Pos.	<i>H</i>	<i>P</i>	<i>L</i>	<i>R</i>	Av. Pos.
<i>A</i>	3	5	6	8	5.5	3	5	5	8	5.2
<i>B</i>	4	8	2	6	5.0	6	8	6	4	6.0
<i>C</i>	2	2	4	1	2.2	2	3	3	2	2.5
<i>D</i>	6	4	5	7	5.5	5	4	4	7	5.0
<i>E</i>	1	1	3	2	1.7	4	2	2	3	2.7
<i>F</i>	7	3	1	3	3.5	1	1	1	1	1.0
<i>G</i>	5	7	7	4	5.7	7	7	7	6	6.7
<i>H</i>	8	6	8	5	7.0	8	6	8	5	7.0

In both cases criteria *F*, *C*, and *E* stand higher than the remaining criteria. But there are nevertheless differences in position among the various criteria which seem sufficient to be significant, viz., the higher positions of *F* and *D* in the case of judgments of "worse." Inference on the basis of specific failures or successes, and direct estimate of total time interval or duration are relied on less when judging "better" than when judging "worse." This means that the direct perceptions of smoothness and of speed are less prominent, as are also feelings of pleasantness and unpleasantness. The judgment of "better," that is to say, is the result of a direct perceptual process. The judgment of "worse" is somewhat more likely to be at least one step removed from direct perception,—to resemble an inference. This seems to mean that the "positive" qualities of smoothness and speed are appreciated immediately and in their own right, while the logically opposite qualities of roughness and slowness are not appreciated in so direct a manner. If this be true, it falls in line, in an interesting way, with previous findings as to the way in which judgments which are logically opposite are psychologically related to each other. Thus two sets of judgments of dislike or of stupidity, in the case of photographs of human faces, show lower correlation than do similar sets of judgments of preference or intelligence, and also yield a higher variability (see Chapter VIII.). Further, before the two categories have been explicitly brought together in the consciousness of the observer, the personal consistency coefficient of two arrangements of given materials on the basis of resemblance to a given stand-

ard is higher than that of two arrangements for unlikeness (see Chapter VII.). Moreover, if an observer is left free to choose the direction of his judgment in comparing two sensory stimuli, there is found to be a strong tendency to direct the judgment toward the stimulus described as "positive" in quality (see Chapter VI.). All of these facts go to show that logical opposites are not necessarily psychological opposites.

TABLE XXIII

COMBINED AVERAGES OF "BETTER" AND "WORSE" JUDGMENTS
Average of 4 Observers for Each Test. Also Number of Cases

Test	A		B		C		D	
	Self	Witness	Self	Witness	Self	Witness	Self	Witness
Col.....	9.8	9.5	5.5	5.2	2.5	3.1	1.3	0.8
Cases....	72	94	128	128	121	103	79	75
Opps....	12.4	10.6	7.5	6.7	3.6	5.3	2.3	4.3
Cases....	108	147	132	110	105	81	55	62
Canc.....	14.2	14.8	7.3	2.3	1.7	3.4	0.6	1.0
Cases....	82	47	115	153	124	118	79	82
Add.....	9.8	11.1	4.3	3.8	2.3	2.1	2.5	2.5
Cases....	112	90	138	133	91	123	59	54
Average..	11.5	11.5	6.1	4.5	2.5	3.5	1.7	2.2
Cases....	374	378	513	524	441	425	272	273
Grand av.	11.5		5.3		3.0		2.0	
Cases....	752		1,037		866		545	

But it will be shown later that this difference in the nature of the criteria is not responsible for the difference in the magnitude of the constant deviations. It will be shown that although the constant deviations are consistently different, they are so related to the per cent. of correct judgments that the probable error (the difference correctly reported in 75 per cent. of the cases) is the same for all circumstances.

2. When the four degrees of confidence are considered, regardless of direction *there is seen to be no appreciable difference between judgments of performer and judgments of witness.* The thresholds are not consistently different, and the distribution of judgments among the various degrees of confidence is almost identical in the two cases (Table X.).

3. *Correctness of Judgment.* If the judgment be classed as right or wrong according as the record on which it was based did or did not depart from the usual performance in the direction indicated in the judgment (regardless of amount) the per cent. of correct judgments may be correlated with the degree of confidence. Table XXIV. summarizes the results of this classification. As in the previous study, correctness increases with certainty, and even pure guesses are more likely to be right than wrong. Roughly, the per cent. cor-

TABLE XXIV
 PER CENT. CORRECT JUDGMENTS

Test	Situation	Better				Worse			
		-A	-B	-C	-D	+A	+B	+C	+D
Col.:	Self.....	80	74	61	53	100	72	59	56
	Cases.....	44	76	71	50	28	52	50	29
	Witness.....	81	71	39	52	92	63	71	50
	Cases.....	80	85	56	35	14	43	47	40
Opps.:	Self.....	92	78	58	42	93	88	72	81
	Cases.....	75	91	62	29	33	41	43	26
	Witness.....	71	63	57	43	96	79	81	74
	Cases.....	129	93	52	29	18	17	29	33
Canc.:	Self.....	96	77	58	63	97	86	81	45
	Cases.....	52	82	68	44	30	33	56	35
	Witness.....	93	69	63	69	100	77	78	64
	Cases.....	36	115	67	46	11	38	51	36
Add.:	Self.....	93	75	67	75	89	68	52	54
	Cases.....	73	86	48	25	39	52	43	34
	Witness.....	95	79	61	92	97	66	59	44
	Cases.....	62	86	58	23	28	47	65	31
Average:	Self.....	90	71	61	58	95	78	66	59
	Cases.....	244	335	249	148	130	178	192	124
Average:	Witness.....	85	70	55	64	96	71	72	58
	Cases.....	307	379	233	133	71	145	192	140
Grand average.....		A = 92		B = 73		C = 63		D = 60	
Total cases.....		686		366		216		332	

rectness (averaging both performer and witness, and both "better" and "worse" judgments) is *A* 90 per cent., *B* 75 per cent., *C* 65 per cent., *D* 60 per cent. In the previous study, in which the three practised observers only were concerned, these percentages were somewhat higher, viz., *A* 98 per cent., *B* 80 per cent., *C* 70 per cent., *D* 60 per cent. Judgments as witness are correct nearly as often, in the long run, as are those of the performer, and, in the case of both, judgments of "better" are somewhat less likely to be correct than are those of "worse" (the average difference being 4 to 5 per cent.).

4. *The threshold variation* in performance for the judgments of all degrees of confidence varies with the general situation in which the judgment is passed. Within each degree of confidence there are four different judgment situations:

- A. Witness judging performer to be worse than usual.
- B. Performer judging self to be worse than usual.
- C. Performer judging self to be better than usual.
- D. Witness judging performer to be better than usual.

The highest threshold is required for situation *A*, then come, in order of diminishing threshold, *B*, *C*, and *D*. Similarly, situations requiring large thresholds show a smaller number of judgments of the given degree of confidence. If it were correct to refer to these facts as the "sensitivity" of the judgments, those judgments being most "sensitive" which require the smallest variations of performance as their basis, the result might be stated as follows:

A. The most "sensitive" judgments are those in which the witness affirms superior performance on the part of another person.

B. Next come the performer's own judgments of himself as "better than usual."

C. Then come the performer's judgments of himself as "worse."

D. Finally, least "sensitive" of all, the witness's judgments of inferiority on the part of another person.

In other words, the thresholds for the witness, as compared with those for the performer, are lower for efficiency and higher for inefficiency. This is also shown in the distribution of the judgments. On the question as to whether these differences indicate genuine differences in "sensitivity" or whether they merely show different judgment attitudes or degrees of predisposition, more will be said later.

5. *Test Differences*.—The four tests may be compared from three different points of view:

A. Average Amount of Variation Required as the Basis for a Judgment of a Given Degree of Confidence.—This comparison may be most easily made by reference to Table XXV. in which the per cent. variation for each degree of confidence is given, the direction of the variation being disregarded and the results of performer and witness being combined.

The test differences are neither considerable nor very consistent—addition, color-naming, cancellation, opposites, is the order in the

TABLE XXV
TEST DIFFERENCES

		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
Colors:	Threshold.....	9.6	5.3	2.8	1.0
	Cases.....	166	256	224	154
Opposites:	Threshold.....	11.5	7.1	4.4	3.3
	Cases.....	255	242	186	117
Cancellation:	Threshold.....	14.5	4.8	2.5	.8
	Cases.....	129	268	242	161
Addition:	Threshold.....	10.4	4.1	2.2	2.5
	Cases.....	202	271	214	113

long run. In the previous study, in which tapping, color-naming, and opposites were used as tests, judgments in color-naming were, as in the present instance, more sensitive than those in opposites, while tapping was twice as sensitive as either of these. It was there suggested that "progressively larger variations in performance are required as the basis for judgments of a given degree of confidence as one passes from an automatic, objectively observable performance (such as tapping), through work involving perceptual reactions (color-naming), to work of a more strictly mental and less objectively observable character (opposites)." No new information on this point is afforded by the present study.

B. Correctness.—In this respect no consistent test differences seem to be present. The lowest per cent. correctness for *A* confidence is found in judgments of "better" in color-naming, but in other cases this test shows up as well as any of the others.

C. Conformity to the Law of Smaller Thresholds for Judgments of "Better."—The chief point to be made here concerns addition. This is the only test in which the law does not hold,—the usual relation of thresholds being found here only in the *B* judgments of the performer and the *A* judgments of the witness. Addition, that is, which is the most sensitive test, shows the law least emphatically. Color-naming and cancellation, which are about equally sensitive, show the law about equally strikingly. Opposites, which is the least sensitive, shows the law most clearly. This seems to mean that the more difficult the judgments (difficulty being measured by the average constant variation required for a given degree of confidence) the stronger is the predisposition toward "better" judgments. Much the same thing was found in the earlier study in which tapping, color-naming, and opposites were compared with each other.

6. Individual Differences.—Tables XX. and XXI. show the individual thresholds when the four tests are averaged. All individuals show the same tendency to pass judgments of "better" on smaller average constant deviations, but they do not show it equally clearly. Observer *L*, whether acting as performer or as witness, always shows the tendency, and under all four degrees of confidence. *H* (the writer) shows the tendency least clearly. *R* and *P* offer occasional exceptions with the lower degrees of confidence. With respect to the magnitude of the variations no consistent individual differences seem to be present.

7. Amount of Variation and Per Cent. Correctness.—In the previous study the figures in the first part of the following table were secured, and in the present study those in the latter part of the table. The $\Delta/P.E.$ for these various percentages of correctness is given as

found by using tables presenting this relation when the per cent. correctness for a given difference is given (Fullerton and Cattell, "Small Differences").

TABLE XXVI

PROBABLE ERRORS

Degree of Confidence <i>A</i>	<i>B</i>		<i>C</i>		<i>D</i>			
1st	2d	1st	2d	1st	2d	1st	2d	
Av. per cent. diff.	10.8	11.5	5.2	5.3	3.2	3.0	1.9	2.0
Per cent. correctness	98	92	81	73	73	63	59	60
Diff. divided by P.E.	3.05	2.08	1.30	.91	.91	.49	.34	.38
Probable error	3.1	5.5	4.0	5.9	3.5	6.1	5.6	5.3
Av. probable error	4.3		4.9		4.8		5.5	

Or if the results of the two experiments be averaged, the following table results:

TABLE XXVII

PROBABLE ERRORS

Degree of Confidence.....	A	B	C	D
Av. per cent. difference.....	11.2	5.3	3.1	2.0
Per cent. correctness.....	95	77	69	60
Diff. divided by P.E.....	2.44	1.10	.69	.38
Probable error.....	4.6	4.8	4.5	5.2

When the probable error is computed in this way, it gives the amount of difference which will be correctly reported 75 per cent. of the times. This P.E. is found to be uniformly about 4.8 per cent. variation from "usual," for all degrees of confidence.

In the same way may be compared the P.E. for "better" and for "worse" judgments. The results are as follows. The table gives the results when the two experiments are combined to give averages.

TABLE XXVIII

PROBABLE ERRORS

Degree of Confidence.....	Judgments of "Better"				Judgments of "Worse"			
	A	B	C	D	A	B	C	D
Av. per cent. difference.....	8.5	4.6	1.9	1.3	13.7	5.8	4.1	1.5
Per cent. correctness.....	90	77	63	58	98	77	72	60
Diff. divided by P.E.....	1.90	1.10	.49	.30	3.05	1.10	.86	.38
Probable error.....	4.5	4.2	4.0	4.3	4.5	5.3	4.3	4.0

The probable error is seen to be, in all cases, about 4.5 per cent. This same P.E. is indicated regardless of the degree of confidence or of the direction of the variation. When the per cent. correctness and the amount of the difference are both taken into account the actual thresholds for judgments of efficiency differ in no way from those of judgments of inefficiency. Reference to the tables shows that when a judgment with a given degree of confidence is passed on the basis of smaller average minus variations than in the case of plus varia-

tions there is usually a falling off in the per cent. correctness. An observer is, then, no more sensitive to gain in efficiency than he is to loss, but he is predisposed to judge both himself and a performer whom he is watching as having done "better than usual" rather than "worse than usual." The consequence is that smaller degrees of superiority tend to be judged as better with higher degrees of confidence, and that a certain slight degree of inferiority tends to be incorrectly judged as "better." It is this situation which is chiefly responsible for the smaller constant variations on which judgments of "better" are based.

If the four different judgment situations be now considered, it will be seen that we were not dealing with genuine differences in "sensitiveness" in the earlier tables. The following table shows that probable error for all four judgment situations is quite the same, the differences in threshold measuring, in reality, not the sensitiveness of judgments but the strength of a predisposition. We are predisposed to judge "better" rather than "worse" and we are, furthermore, predisposed in favor of the other man rather than of ourselves.

TABLE XXIX
JUDGMENT SITUATIONS

Situation	Degree of Confidence	A	B	C	D
Witness judging performer to be "worse than usual":	Av. per cent. difference.	14.5	5.0	4.9	1.7
	Per cent. correct.	96	71	72	58
	Diff. div. by P.E.	2.60	.82	.86	.30
	Probable error.	5.6	6.1	5.7	5.6
	Av. P.E., disregarding D judgments.		5.8		
Performer judging self to be "worse than usual":	Av. per cent. diff.	13.8	7.7	3.2	.2
	Per cent. correct.	95	78	66	59
	Diff. div. by P.E.	2.44	1.14	.61	.34
	Probable error.	5.6	6.7	5.3	.6
	Av. P.E., disregarding D judgments.		5.8		
Performer judging self to be "better than usual":	Av. per cent. difference.	9.2	4.7	2.6	1.3
	Per cent. correct.	90	71	61	58
	Diff. div. by P.E.	1.90	.82	.41	.30
	Probable error.	4.8	5.7	6.3	4.3
	Av. P.E., disregarding D judgments.		5.6		
Witness judging performer to be "better than usual":	Av. per cent. difference.	8.1	3.9	1.3	1.0
	Per cent. correct.	85	70	55	64
	Diff. div. by P.E.	1.54	.78	.19	.53
	Probable error.	5.3	5.0	6.8	2.0
	Av. P.E., disregarding D judgments.		5.7		

The differences found do not then indicate real differences in sensitivity under the various judgment situations,—they measure the relative strength of these various predispositions, tendencies, and inclinations. These observers were, under all circumstances, disinclined to judge any trial as “worse than usual,” and the disinclination was stronger when judging as witness than when judging as performer. This results in a combination of optimism and altruism which, if found to be a common occurrence, would seem to have exceedingly interesting psychological and perhaps social implication. Further investigation will perhaps show that these predispositions are conditioned, under different circumstances, by a variety of factors, such as competition, education, motive, age, sex of performer and witness, and perhaps by individual differences of a temperamental sort.

CHAPTER IV

THE CENTRAL TENDENCY OF JUDGMENT¹

SINCE the work of the early investigators of the time sense the concept of the "indifference point" (I.P.) has played an ever-present rôle in experiments on judgments of magnitude, duration, and intensity. Judgments of time, weight, force, brightness, extent of movement, length, area, size of angles, have all shown the same tendency to gravitate toward a mean magnitude, the result being that stimuli above that point in the objective scale were underestimated and stimuli below overestimated, while the mean magnitude itself was invested with no constant error. This region in the scale, flanked above and below by negative and positive constant errors, was called the indifference point, or more properly the region of indifference.

The tendency has been throughout to infer that the I.P. disclosed in any particular experiment was in some way an absolute quantity and should be found in other experiments on the same quality of stimulus. In this way arose the ideas of a "most favorable extent" (Kramer and Moskiewicz, Jaensch) and a "most favorable time" (Vierordt, Höring, Estel, etc.). Among the investigators of the time sense, since an I.P. was found for every group of intervals employed, grew up the doctrine of periodic I.P.'s, those for regions higher up in the scale being multiples of the I.P.'s found in the experiment in which the shortest intervals were used. Attempts were made to correlate the unit of periodicity with various bodily processes—the swing of the leg, breathing time, pulse beat (Wundt, Münsterberg). All of this speculation passed the criticism of laboratory workers and was incorporated in the general texts as a curious fact, productive of many illusions and constant errors, but the analysis was carried no farther.

In an earlier study² the writer undertook an experimental analysis of the phenomenon of the I.P. in judgments of the duration and extent of rectilinear arm movements. The results of this investigation showed conclusively that, with the method of reproduction, the following principles hold.

¹ Reprinted from *The Journal of Philosophy, Psychology, and Scientific Methods*, Vol. VII., No. 17, August 18, 1910.

² "The Inaccuracy of Movement," H. L. Hollingworth, *Columbia Contributions*, Vol. XVII., No. 3, June, 1909.

I. The I.P. is relative,—not absolute. It is a function of the series limits of the stimuli employed. Given the series of magnitudes with which we are to work, we may be quite certain that a region of indifference will occur at about the midpoint of that particular scale.

II. A periodic I.P. can be found within a total series (*S*) by working with its special sections (*A*, *B*, and *C*).

III. The same absolute magnitude may be either an I.P., or affected with a positive constant error, or with a negative constant error, according to the particular range or section in which it occurs.

IV. The gradual extension of the series limits is accompanied by a corresponding shift in the region of indifference.

V. No magnitude estimated out of relation to a series or group of which it is a member evinces any considerable constant error.

VI. The phenomenon of the I.P. disappears as the interval between separate judgments is extended. The first disposition is soon dissipated and is no longer adequate to affect the second performance.

VII. In a parallel tabulation of the I.P.'s and the ranges of intervals used in the various time-sense studies the influence of the latter on the magnitude of the I.P. is clearly seen.

VIII. The phenomenon of the I.P. and the so-called positive and negative time errors result from a general law—the *central tendency of judgment*. In all estimates of stimuli belonging to a given range or group we tend to form our judgments around the median value of the series—toward this mean each judgment is shifted by virtue of a mental set corresponding to the particular range in question. This *central tendency* is not a “law of sense memory.” It is a law of immediate perception and disappears as the experiment becomes a memory test.

IX. In experiments by the method of reproduction this *central tendency* is reenforced by the law of motor habit.

For an account of the experiments on which these conclusions rest and for detailed exposition of their significance the reader must be referred to the earlier study.

The Present Study

Purpose.—On account of the reenforcing value of the law of motor habit the earlier experiments did not indicate how clearly or in how far the results secured were a function of the method of motor reproduction. In order to support the case completely it should be shown that the same law of judgment is present in experiments into which the method of reproduction does not enter.

In order to put the generalization to such a test the following experiments have been made on judgments of the size of squares, by the method of selection.

Observers.—The observers were all women students in Barnard College with from one and a half to two and a half years of training in psychology. Different observers were used in the two experiments and none of them knew the purpose of the experiment, nor were they familiar with the results of the earlier study.

Material.—The material used in both experiments *A* and *B* was the same, the chief differences between the experiments consisting in the way in which the series limits were varied. On a dark gray wall were placed 30 squares of light gray cardboard, ranging in size from 2.5 cm. on a side to 50 cm. and increasing from 2.5 to 7 cm. by increments of 0.5 cm., from 7 to 15 cm. by increments of 1 cm., from 15 to 40 cm. by increments of 2.5 cm., and on to 50 cm. by increments of 5 cm. Each card was numbered in consecutive order. Alongside these standard cards and at the same distance from the observer was an exposure apparatus, by means of which, at proper intervals, the fourteen test cards could be presented one at a time. These test cards varied in size from 3 cm. to 40 cm. on the side, ranging from 3 to 7 cm. by increments of 1 cm., from 7 to 15 cm. by increments of 2 cm., from 15 to 40 cm. by increments of 5 cm.

Procedure.—In each experiment a test card was exposed for 5 seconds. The observer then waited for 5 seconds, the eyes resting meanwhile on a dark screen. She then turned to the standard series and was allowed 5 seconds in which to select a card corresponding in size to the one just exposed and to write its number in her record. A second test card was then exposed, and so on throughout the experiment. By keeping a record of the order in which the test cards were shown, the experimenter was able subsequently to compare the observer's judgment with the actual magnitude. As a result of this method of selection all constant errors due to the law of motor habit in reproduction are eliminated and any error disclosed will be entirely an error of judgment of visual magnitude.

Experiment A

This experiment began with series 3, 4, 5, 6, 7, three trials for each magnitude, in chance order. The smallest card (3) was then dropped and the larger card (9) substituted, and three trials taken in chance order, for each member in the new series 4, 5, 6, 7, 9. In this way the successive series moved up along the total range, dropping at each change the lowest member and including the one next

larger than the greatest number. The series, that is to say, always consisted of 5 test cards, and as the experiment progressed, magnitudes were dropped from the lower end and new ones added to the upper end. Ten observers were used, 150 trials being taken on each observer. Table XXX. gives the C.E. of the 10 observers in terms of the square root of the area—that is, in terms of the length of one side of the square. Each figure is the C.E. resulting from 30 judgments.

TABLE XXX

GIVES THE C.E. IN CM. OF EACH CARD IN EXPERIMENT A.

10 OBSERVERS, 1,500 TRIALS

	3	4	5	6	7	9	11	13	15	20	25	30	35	40
1	0	-.13	-.23	-.24	-.21									
2		+.15	+.52	+.53	-.01	+.44								
3			+.51	+.15	-.11	+.32	+.31							
4				+.19	+.39	+.55	+.21	-.02						
5					+.31	+.21	+.42	-.13	0					
6						+.74	+.75	+.64	+.56	+.48				
7							+.131	+.80	+.137	+.173	+.215			
8								+.139	+.160	+.184	+.143	+.192		
9									+.94	+.172	+.215	+.98	+.90	
10										+.240	+.265	+.150	+.45	+.178

Experiment B

This experiment began with the series 3, 4, 5, 6, 7, 9. Three trials for each magnitude were taken in chance order. The next higher magnitude (11) was then added to the series and again 3 trials for each magnitude (3-11) were taken in chance order. At this point the next magnitude (13) was introduced, 3 trials for each card taken, and the process continued until in the ninth series the whole range of test cards from 3 to 40 was included. Six observers were used, 270 records being taken from each observer. Table XXXI. gives the C.E. of the 6 observers for each magnitude in each suc-

TABLE XXXI

GIVES THE C.E. OF EACH CARD IN EXPERIMENT B.

6 OBSERVERS, 1,620 TRIALS

	3	4	5	6	7	9	11	13	15	20	25	30	35	40
1	.03	.10	.08	.42	.25	.58								
2	.03	.17	.15	.45	.25	.65	.86							
3	.03	.26	.48	.60	.11	.80	.89	.60						
4	.03	.53	.73	.88	.45	.40	.53	.65	1.43					
5	.03	.65	.98	.83	1.05	.43	.36	.52	1.60	2.63				
6	.05	.65	1.05	.78	.85	.72	.43	-.25	1.62	2.05	2.40			
7	.03	.76	1.05	.90	.92	.93	.80	1.00	1.35	1.73	2.25	4.82		
8	.05	.87	1.12	.73	1.23	.70	.82	1.83	1.27	1.77	1.63	1.85	3.08	
9	.08	.68	1.08	.87	1.10	.75	.42	.92	1.52	1.57	1.43	.97	2.10	4.42

ceeding series. As in Table XXX. the errors are given in terms of one side of the square. Each figure in the table is the C.E. of 18 judgments of the same card.

In each of these experiments we have another case of the gradual extension of series limits, and if the *law of central tendency* is operative, I.P.'s might be expected to occur in each series and gradually to rise in the range as the larger magnitudes are added. The A.E. and its variability are not given in the tables, since only the C.E. is of interest for the problem in hand. As a matter of fact the phenomenon of the I.P. is concealed in both experiments by a strong positive constant error which comes from a general tendency to overestimation in judgments of square magnitudes. This tendency has been found by other investigators. Woodworth and Thorndike find a positive constant error in estimates of area by a mental standard. Baldwin, Shaw, and Warren find the same tendency in judgments of the size of squares and attribute it to a change in the memory image. This error, however, is irrelevant to the present problem. The important fact is that underneath this ever-present overestimation the *law of central tendency* is also operative, and its presence can be clearly shown by a proper analysis of the figures.

Casual examination of Table XXX. shows that the positive constant error for any one magnitude increases as the place of the magnitude in the series descends. Thus the — C.E. (— .21) for card 7 in series 1 changes to a decided + C.E. (+ .39, + .31) in series 4 and 5. The + C.E. (+ .31) of card 11 increases to + 1.31 in series 7, and the errors of the other cards undergo in a strikingly uniform way the same transformation. This is a clear indication that in any one series the magnitude is influenced by other magnitudes occurring above and below it and is in every case shifted toward the center of the series. Thus in series 1 card 7 is drawn toward the smaller magnitudes, and its judgment results in a — C.E. In series 5 the same card is drawn toward a higher set of magnitudes and hence acquires a decided + C.E.

The process is clearly shown by an examination of the 6 cards (7 to 20, inclusive) that occurred in all 10 series. Each of these cards occupied, in the course of the experiment, all 5 positions. Thus card 11. is in series 3 the largest magnitude; in series 7 it is the lowest; in series 5 it is the central card; while in series 4 and 6 it occupies the intermediate positions on either side of the center. The same, in appropriate series, is true of all 6 cards, from 7 to 20 inclusive. Now if there were no source of error present except the *central tendency of judgment* each card should have theoretically

no C.E. when it occurred in the middle of a series, *i. e.*, it should be the I.P. for that series. But, since there is another error present due to the general tendency to overestimation in judgments of square size, the theoretical conditions are not fulfilled, and each card, even when it occurs in the central position, shows an actual $+$ C.E. We may assume, then, that the error shown in this central position is due to the character of the material, and that so far as the *law of central tendency* is concerned it may be considered 0, or what we might call the *normal error*. If the errors of any magnitude in the successive series from 1-10 be calculated with respect to this *normal error*, the operation of the *law of central tendency* should lead to the following results. As the series progress the relative errors of any magnitude, that is, the deviations of the actual from the *normal errors*, should show an I.P. phenomenon—they should be negative above the normal, zero at the normal, and positive below it. The facts are shown in Table XXXII., in which, for cards 7-20, the error of each card when it occurred in central position is assumed to be *normal*. It will be seen that above the normal the errors are, with a single exception, negative, while below they are, with only three exceptions, positive. The transformation is from a high $-$ value through 0 to a high $+$ value.

TABLE XXXII

7	9	11	13	15	20
-.10	-.11	-.11	-.66	-1.37	-1.32
+.10	-.23	-.21	-.77	-.81	-.11
0	0	0	0	0	0
+.28	-.34	+.33	+.16	+.23	-.12
+.20	+.19	+.89	+.75	+.43	+.68

Thus from any point of view in which the figures may be regarded the *central tendency of judgment* is revealed, working, however, underneath a general tendency to overestimation. This result is confirmed by the results of Experiment B, in which the lower magnitudes were allowed to remain in the series while the higher were being added. The results appear in Table XXXI. Again there is present the positive constant error due to the character of the material, but underneath the *central tendency* is clearly to be seen.

The magnitudes here used fall into three groups. To the first group belong cards 3-9, present in all 9 series, and influenced in judgment by the gradual inclusion of the higher magnitudes 11-40. According to the aforestated law the effect of these higher magnitudes should be to draw the lower cards toward a constantly augmenting center, that is, as the higher cards appear one by one, the

central tendency of the respective series rises. The positive errors of cards 3-9 should thus become constantly greater as the experiment proceeds. Again the deductions are strikingly verified. Thus the error of card 4 increases from $+.10$ in series 1 to $+.68$ in series 9; that of 5 from $+.08$ in series 1 to $+1.08$ in series 9, etc. This effect is due, in any one series, partly to the introduction of still higher magnitudes, partly to habituation to the larger cards already introduced and now being repeated.

The second group of magnitudes consists of cards 20 to 40 inclusive. When any one of these cards, say 20, is introduced, the observer is already considerably adapted to the lower magnitudes, and as the next higher card (25) is introduced in the following series this adaptation to the lower cards is much furthered by the fact that each of the 9 cards below 20 is again repeated three times, while adaptation to magnitudes higher than 20 is only slightly begun by the threefold repetition of card 25. The consequence is that as the experiment proceeds habituation to the lower range increases much more rapidly, at first, than that to the upper range, on account of the greater number of lower cards. In this group, then, we should expect transformations just the reverse of those in group I., that is, the $+C.E.$'s should become constantly smaller as the high card is drawn more and more in judgment toward the center of the series. Again expectation is confirmed. The error of card 20 falls from $+2.63$ in series 5 to $+1.57$ in series 9; that of card 25 from $+2.40$ to $+1.43$; that of card 30 from $+4.82$ to $+.97$; and that of card 35 from $+3.08$ to $+2.10$.

There remain yet to be considered the three cards 11, 13, and 15, comprising group three. This group, standing as it does midway between groups one and two, which show directly opposite transformations, might be expected to show either of two results. First, the two tendencies might neutralize each other, the errors in group three remaining approximately constant or varying irregularly. Second, the first tendency might operate in the first few series, after which, by virtue of increasing habituation to the larger cards the second tendency might begin to assert itself in the later series. So far as the figures go they are sufficiently irregular to admit of either interpretation. There is neither uniform increase nor decrease throughout. There is, in fact, a strong suggestion of the second possible result—initial decrease followed by increase as habituation to higher magnitudes grows. Thus the errors of card 11 fall from $+.86$ in series 2 to $+.36$ in series 5, then increase to $+.80$ and $+.82$ in later series. Card 13 falls from $+.60$ in series 1 to $-.25$ in series 6, then increases to over $+1.00$ in series 7-9. Card

TABLE XXXIII

	3	4	5	6	7	9
1	-.01	-.42	-.67	-.30	-.44	-.08
2	-.01	-.35	-.60	-.27	-.44	-.01
3	-.01	-.26	-.27	-.12	-.58	+.14
4	-.01	+.01	-.02	+.16	-.24	-.26
5	-.01	+.13	+.23	+.11	+.36	-.23
6	+.01	+.13	+.30	+.06	+.16	+.06
7	-.01	+.24	+.30	+.18	+.23	+.27
8	+.01	+.35	+.37	+.01	+.59	+.04
9	+.04	+.16	+.33	+.15	+.41	+.09

15 falls to +1.35 in series 7, increasing to +1.50 in the last series.

One could scarcely ask for more convincing evidence of the *law of central tendency* than that afforded by the behavior of the C.E.'s in these three groups of magnitudes. The evidence may be re-enforced, however, and the process more clearly exhibited by further treatment of the errors in group I., consisting of cards which were present in all 9 series. In the case of this experiment we have no means of determining, as we did in experiment A, the normal error due to the character of the material. We may, however, observe the deviations of the errors in a given series from the average of the errors in the whole 9 series. These deviations should show, as did Table XXXII. for experiment A, an indifference point phenomenon for the errors of any given magnitude in successive series. Such a calculation results in Table XXXIII. As was to be expected, the I.P. phenomenon is clearly present. The successive deviations from the average, in the case of the errors for any given magnitude, pass from pronounced negative direction through an approximate zero point to a pronounced positive direction. This change was caused in every case by the inclusion of higher magnitudes in the series, thus producing an upward shift in the central tendency or median of the series, toward which each lower magnitude was assimilated in greater or less degree, according to the amount of habituation to the upper range.

It is not necessary to go further into the theoretical and interpretative consideration of the *law of central tendency*, since the writer has already discussed this elsewhere.³ But it should be pointed out that none of the factors usually introduced to explain the occurrence of indifference points are adequate. Unexplained differences in time error (Fechner), mechanical sources of error in apparatus (Schumann), peculiarity of the sense organ (Vierordt), lack of current motor control (Delabarre), relative expenditure of energy (Wundt), change in the memory image (Wreschner, Leuba),

³ "Inaccuracy of Movement," Chapter III.

fatigue and dynamogeny, all these may contribute their share toward the actual magnitude of a given error, but their influence can hardly be conceived as varying up and down a scale of objective magnitudes in such a way as to account for the shifting I.P. with extension of the series limits.

Nor is the phenomenon in any way the result of contrast. It is, on the contrary, just the reverse—a case of two magnitudes approximating each other in judgment by virtue of their temporal contiguity. The tendency seems explicable only in terms of itself. Just as our experience with a race, class, or social group results in the conception of a *type* which shall in some way represent the central tendency of the group, and from which the separate members shall deviate the least, so in an experiment on sensible discrimination we become adapted to the median value of the series, tend to expect it, to assimilate all other values toward it, and to greater or less degree to substitute it for them. Either this tendency is the rudimentary process out of which the higher acts of conception grow, or it is the habit of conception extended to sensory fields and interfering with a quite elementary process of comparison and recognition. The importance of the law in any series of psychophysical measurements should be apparent. The error to which it leads is distinctly an error of judgment, and is quite independent of sensory or physiological conditions which may of themselves be sources of other types of errors.

CHAPTER V

THE DIRECTION OF JUDGMENT

So far as the writer is aware the only discussion of the influence of the direction of judgment is to be found in the works on psychophysics. In these works the problem is handled chiefly as a point in experimental technique and treated as an issue which must be disposed of before some further problem can be most precisely approached. In the several papers that follow this chapter the phenomena of preferred or accustomed directions, inclinations, or tendencies of judgment, and the influence, on the outcome of the judgment, of the form or category in which it is expressed, are themselves to have the place of chief interest. In place of the simple stimuli used in the psycho-physical studies, material of a more complex sort has been employed. This has been done partly because of immediate interest in these little-studied subjective types of judgment, and partly because of a preliminary assumption that this kind of material would involve processes and criteria which might be more sensitive to the influences just mentioned than might be the case with descriptively simpler and more objectively measurable material.

By way of introduction to the three chapters which follow it may be of interest to sketch briefly some of the chief sections in the literature of psycho-physics in which the problem of the direction of judgment has been raised.

In Fechner's experiments on the discrimination of weights the observer was required to pass one of two kinds of judgments,—he might designate the *heavier* weight or he might express himself as *uncertain*. When a comparison was expressed, that is to say, the direction of judgment was determined by the quality of the stimulus rather than by its time or space order. The subject of the proposition expressing the judgment was always the *heavier* weight, which might be either the right or left, the first or second, in order of presentation.

G. E. Müller¹ devotes considerable space to a preliminary discussion of "die Urtheilsrichtung." Müller points out that which of the six possible ways of expressing the relation between a standard and a variable stimulus is used (indicating the heaviest or lightest, or

¹ "Die Gesichtspunkte und die Tatsachen der psychophysischen Methodik," p. 16 ff.

describing the first or second, right or left) is not a matter of indifference. For at least three reasons the observer should always, on beginning an experiment, be given definite instructions with respect to the direction of his judgments, and these instructions should be recorded. In the first place the six directions differ in convenience and ease, both for operator and for observer. In the second place, the results of some methods of instruction are more informative than others. Finally the part played by "absolute impression" depends somewhat on the direction of attention toward the one or the other stimulus. These remarks hold whether the order of presentation be simultaneous or successive.

The instruction to direct the judgment always toward the standard or toward the variable, Müller dismisses because of the danger of confusion, either in the mind of the observer or in the records. Nor is the method of periodically changing direction felt to be satisfactory. When the two stimuli are simultaneous the preferable procedure is held to be that of "free direction" in which, whether the judgment shall relate to the first or second, standard or variable, heavier or lighter, is left to the option of the observer. Two reasons are given for this preference for the method of "free direction." The first is found in the statement that such procedure "does least possible violence to the psychological tendency of the observer." The second is the fact that, given a good observer and an appropriately planned experiment, information can be secured concerning the observer's type and his attention characteristics by examining the frequency of the various forms of judgment. It was by utilizing this method that Müller classified his observers as positive or negative in type.

In the case of successive stimuli Müller believes that the method in which the judgment always relates to the second stimulus is far superior to any other method of "prescribed direction,"—"because this is the *simplest, most natural* method, and the one most free from omissions and confusions." No experiments with successive presentation and free direction of judgment are recorded. Müller however asserts that the method of "free direction" with respect to space position is always to be recommended. With respect to temporal position no experiments are recorded. The same is true of procedure with *absolutely free* direction in which judgment may refer, at the discretion of the observer, toward either the right or left, first or second, stimulus.

Three points are to be noted in Müller's discussion. One is the statement that if the direction of judgment is to be prescribed, the direction should always be *toward the second stimulus* because this is the "*simplest and most natural* method." The second is the state-

ment that observers have *psychological tendencies* which may be violated. The third is the assertion that the *direction of attention* may influence the distribution of the judgments.

Müller and Schumann instructed their observers to direct their judgment *toward the second stimulus* presented. Martin and Müller (*Unterschiedsempfindlichkeit*) experimented by various methods, such as judgment on the (a) variable, (b) standard, (c) heavier, (d) second, ignoring the method of judging always on the first. Fechner's method (judging which is heavier) is said to complicate unduly the process of judgment. It is asserted that if the difference between standard and variable is clear the observer always decides at once how the second compares with the first, and the reply is made much more easily under the Müller and Schumann method (judgment on the second). "If the observer must say which is lighter or which is heavier the psychological process is too complex. Subjects complain of having to hold the impression in memory while deciding its position." This method is also objected to because of difficulties in recording, on the part of the operator.

Methods with judgment always on standard or on variable are also reported to be both unnatural and too complex, and to present difficulties in the matter of records. This leaves the Müller and Schumann method as the preferable procedure. But it should not fail to be noticed that the "difficulties" and "complexities" of the other methods are for the most part, reported by the operator, or on the part of observers already long practised in the Müller and Schumann method.

Fullerton and Cattell² in experiments on extent of movement, on lifted weights and on lights, instructed their observers to state the relation of the second to the first stimulus. In the general discussion of the psycho-physical methods these investigators state that "the method of right and wrong cases—in which two stimuli nearly alike are presented to an observer and he is required to say *which seems the greater*—is the most accurate method" (p. 150). But this seems, in the light of their procedure, to have meant not that the *category of greatness* should be employed, but that the magnitude or intensity of the second be compared with that of the first. The second stimulus was always the subject of the proposition expressing the judgment. Fullerton and Cattell do not take up the question of "direction of judgment" for its own sake.

Titchener³ in describing the method of right and wrong cases advises that "*O judges always in terms of the weight lifted second,*"

² "Small Differences."

³ "Experimental Psychology, Student's Quantitative Manual," p. 119.

and refers, by way of reasons for this procedure, to Müller's discussion.

Warner Brown, in an interesting study of the various factors influencing the judgment of difference in the case of lifted weights⁴ compared, with one observer, the Fechnerian method with that of Müller and Schumann. In discussing the difference between them Brown remarks on the way in which the form of expression may, by inducing a particular mental set or bias, modify the total distribution of the judgments. The following paragraphs are quoted.

"The group which appears to better advantage here is that which adopts the procedure recommended by Müller and Schumann. It has less errors in all and a less dispersion of errors toward the larger differences. It also shows a less exaggerated constant error. So far as the small number of cases warrants any conclusion, it seems also to present a more symmetrical distribution of plus and minus errors, and to have greater regularity. . . . The results leave no doubt that a difference in the framing of two propositions which are precisely equivalent logically will be a governing factor in making a comparison. Evidently no comparison is complete with the mere apprehension of the presented stimuli. These are apprehended in the light of other stimuli which have gone before, but even then the analysis is not complete without taking account of what the observer has to do in the matter. Even the slightest differences in the task which he has to perform seem to govern to some extent his decisions."

"To speak of the 'perception of difference' in such a case is to obscure some of the factors in the actual situation. The difference is not merely perceived. The process of comparison involves the active operation of the mind in the expression of a judgment upon the situation in which the difference is only one factor. When this difference is acted upon through one set of categories and with one mental set it occasions one definite reaction, while if it is taken into another set of categories it goes through different mental machinery and comes out different. If it were possible to catch an instantaneous view of the two experimental groups under consideration, there is no doubt that a weight of 95.5 grams would be sensibly lighter than 100 in the one and heavier in the other. The stimuli to be compared are identical and the difference involved is not conceivably other than identical. Moreover the logical relations of the terms are equivalent. And yet this difference comes out plus in one group and minus in the other. In the instantaneous view it is judged to be sensibly other; to be two distinct differences."

". . . If it be true that the mind will more readily give expres-

⁴"The Judgment of Difference," *California Studies in Psychology*, No. 1.

sion to 'greater' than to 'less,' the fault is certainly not in the perception of the particular difference but rather in the mind's attitude toward all differences. Such a defect would permeate all quantitative judgments and would, in fact, be a defect of judgment itself. There seems to be evidence that some of the abnormalities observed in the comparison of weights are traceable to such subtle eccentricities in the machinery by which all judgments of difference, in any material, are expressed."

Henmon has recently reported observation of decided preferences in the direction of judgments of length of lines.⁵ "One curious constant error in judgments of the shorter line appeared in the results. All of the subjects, particularly *Br* and *H*, noted early in the experiments that judgments could be more easily given, more quickly, and with greater confidence when reaction was to be made to the shorter line. The feeling that the most accurate judgments would be secured with the shorter line was very marked. . . . The results in part confirm the introspections and in part do not. The general averages show in each case that the greater number of wrong judgments was obtained to the shorter line though the differences are not significant except in the case of *Br*. However the number of right *A* judgments (judgments with high degree of assurance) to the shorter line is almost twice as great as to the longer line, except in the case of *Bl* where the difference is not marked."

Burt⁶ remarks: "It may be of interest to note, as bearing on the psychological theory of comparison of sense impressions, that the natural tendency of the boys seemed invariably to be indicative, by pointing or naming, *the heavier* of the two weights, rather than to pronounce a judgment directly expressing an 'absolute impression of the heaviness or lightness of the last lifted.'"

The Present Studies

In the three following chapters, on "Natural or Habitual Tendencies of Judgment," "Judgments of Similarity and Difference," and "The Influence of Form and Category on the Outcome of a Judgment," will be reported a series of experimental inquiries designed chiefly to discover the character and degree of such natural or habitual tendencies or inclinations of judgment as are revealed under experimental conditions, to investigate any individual differences that may be indicated, and to examine into the way in which changes in logical category or form of expression may influence the outcome,

⁵ "Time and Accuracy of Judgment," *Psych. Rev.*, May, 1911, p. 193.

⁶ "Experimental Tests of General Intelligence," *Brit. Jour. Psychol.*, 1909, p. 20.

the consistency, and the variability of judgment. Special attention will be given to the psychological process and criteria underlying judgments which are, from a grammatical or logical point of view, only two sides or modes of expression of one and the same intellectual act. The interest throughout will not be in technique of experimental procedure as has been the case for the most part in the studies just referred to, nor will any attention be given to the relation between objective measurement and subjective estimation. The interest will be in the judgments themselves, their behavior and criteria, and the way in which these are influenced by changes in the task, situation, or mental set in the interest of which the judgment is passed.

CHAPTER VI

NATURAL OR HABITUAL TENDENCIES OF JUDGMENT¹

THE preceding studies have demonstrated the important part played by direction, form, and category in determining the outcome, consistency, and variability of judgment. The present study reports an attempt to learn whether there are some tendencies, categories, or forms of expression which are most naturally or habitually employed, and to learn how such inclinations, if present, vary with individual, with age, and with the modality or general situation in which the judgment is passed. The experiments have been performed on naïve subjects, who neither knew the purpose of the experiments nor were practised in any of the psycho-physical methods. They are moreover limited to results from a group of school children and a group of college students (women). The original plan included a group of male observers but the conditions under which the work was done have made it impossible to secure this third group of observers. The original plan included also a study of the way in which the preferred direction of judgment might vary with the position of the group of stimuli in the total possible range of magnitudes, intensities, etc. But this first section (here reported) proved to require a longer time for its completion than had been expected. Unavoidable interruptions also occurred, so that by the time it was finished the same observers and assistance were no longer available. These further questions, although not discussed in this paper, seem to constitute extremely interesting topics of research and it is hoped that on some later occasion or by some other investigator they may be taken up anew. The method and procedure are here described in detail in order that such later work may be planned on a comparable basis.

The Method of the Experiment

Fifteen sets of stimuli were provided, so chosen as (1) to enable the study of several modalities of sensation, (2) to call for a variety of typical kinds of judgment categories, and (3) to afford, in each set, three degrees of difference, all of which should, however, be easily perceptible. The stimuli used, and their measurements or quality, are here listed.

¹ This experiment was conducted, under the writer's general supervision, by Miss M. E. Bishop, who is also responsible for the tabulation of the data.

Three weights, weighing respectively 25, 40, and 70 grams.

Three heavily drawn horizontal lines, 6, 7, and 8.7 cm. in length.

Cards bearing squares, the sides being 1.5, 2, and 2.5 cm.

Balls of rubber, three different sizes.

Three tuning forks, pitch C, E, and G.

Tones on monochord, lengths of string, 50, 60, and 70 cm. String constant.

Three shades of gray paper, easily discriminable.

Cards bearing in figures amounts of money, \$197.35, \$205.72, and \$628.43.

A pain point (thorn) applied with three degrees of force.

Bottles of violet perfume, two strengths, and a bottle of clear water.

Cards bearing following dates: 1492, 1609, 1776.

Hard rubber ball, falling on floor from heights of 1, 2, and 3 ft.

Three sheets of sand paper, of different degrees of roughness.

Metronome beating at three rates, 76, 100, and 126.

Three wrapped bottles, two containing old cheese of different strength, the remaining bottle containing only water.

These stimuli were presented to 31 observers (21 adults, for the most part students in Barnard College or teachers, all women) and 10 children in the Speyer School (5 boys and 5 girls, ages 11 or 12). In each case two of the stimuli from a given group were given in succession, with an interval of a few seconds. Six trials were made within each group of stimuli, thus giving a total of 90 judgments for each of the 31 observers,—in all 2,790 judgments. Three of these six trials were what will be designated as “positive first, followed by negative.” The remaining three were “negative first, followed by positive.” The use of the terms “positive” and “negative” in this connection is chiefly a matter of convenience. By a negative stimulus is meant simply a stimulus which presents a smaller amount or degree of that quality, force, or property, etc., which characterizes the group. Thus

The observer was requested to compare the two stimuli with respect to some category which was more general than either the positive or negative quality, care being taken not to suggest either the one or the other quality or form of expression. Thus, “Compare these two tones in *pitch*,” “Compare these two squares as to *size*,” “these two odors, as to how they *affect* you,” etc., etc. In the case of the grays, the surfaces, and the lines, however, it was not so easy to give a general instruction which should not more or less directly suggest one or other of the forms of expression available for the judgment. In these cases the observer was simply asked to compare

In judging volumes (balls)	"Positive" means.....larger.
pitches (forks)	"Positive" means.....higher.
shades of gray	"Positive" means.....darker.
amounts of money	"Positive" means.....greater.
pains (prick of point)	"Positive" means.....more acute.
perfumes (violet)	"Positive" means.....more agreeable.
stinks (cheese)	"Positive" means.....more agreeable.
dates	"Positive" means.....later.
weights (pressure)	"Positive" means.....heavier.
sounds (intensity)	"Positive" means.....louder.
surfaces (sandpaper)	"Positive" means.....rougher.
speeds (metronome)	"Positive" means.....faster.
weights (lifted)	"Positive" means.....heavier.
lines	"Positive" means.....longer.
squares	"Positive" means.....larger.

the two. If he hit upon the right comparison, the experiment was continued without further instruction for that group. If the comparison was not of the type desired, he was asked to compare them in still another respect. When the desired comparison was once made, he was asked to compare the remaining stimuli of the group.

That is to say, the observer was left free to select both the direction of the judgment (as to first or second stimulus) and the form of expression (positive or negative quality). This was of course the whole point of the experiment, and the question of interest was: when an observer is left thus free, both as to direction and as to category, what is the direction or form which his judgment most naturally or habitually takes? Does he show any inclination to judge the character of the second stimulus rather than that of the first, or is the direction determined perhaps by some more or less constant tendency to attend to the stimulus possessing either the positive or the negative quality or degree of quality? If, to the naïve observer one direction or one category is either more natural, more accustomed or more easily employed, and if individuals differ in these respects, when the differences between stimuli are clear, the records of 90 judgments by each individual, in the various modalities or types of comparison, ought to disclose the tendencies.

Record was made, in each case, of the order in which the two stimuli were presented, and the stimulus indicated which became the subject of the proposition expressing the judgment. This record enables a statement of the number of judgments directed toward the first or the second, and toward the positive or negative stimulus. The various arrangements were presented in a chance order, care being taken only that the same number of each arrangement be presented,—three of each in each group of six.

In Tables I. and II. the distribution of the 90 judgments, for each

observer, regardless of modality or situation, is given. The records for "positive stimulus first" are kept separated from those for "negative first," but the total distribution also given. It would suffice to give in the table only a statement as to whether the judgment was directed in each case toward the first or toward the second stimulus,

TABLE XXXIV

DISTRIBUTION OF JUDGMENTS. TEACHERS AND COLLEGE STUDENTS

Observer	Positive Quality First				Negative Quality First				Total Distribution			
	1st	2d	Pos.	Neg.	1st	2d	Pos.	Neg.	1st	2d	Pos.	Neg.
Lst.....	32	13	32	13	17	28	28	17	49	41	60	30
Ger.....	36	9	36	9	9	36	36	9	45	45	72	18
Stf.....	42	3	42	3	2	43	43	2	44	46	85	5
Bro.....	16	29	16	29	5	40	40	5	21	69	56	34
Mes.....	38	7	38	7	3	42	42	3	41	49	80	10
Sch.....	41	4	41	4	4	41	41	4	45	45	82	8
Schl.....	39	6	39	6	6	39	39	6	45	45	78	12
Sal.....	42	3	42	3	3	42	42	3	45	45	84	6
Bok.....	42	3	42	3	4	41	41	4	46	44	83	7
Mor.....	36	9	36	9	0	45	45	0	36	54	81	9
Ell.....	33	12	33	12	9	36	36	9	42	48	69	21
New.....	39	6	39	6	3	42	42	3	42	48	81	9
Seb.....	22	23	22	23	3	42	42	3	25	65	64	26
Hrt.....	39	6	39	6	6	39	39	6	45	45	78	12
Sav.....	41	4	41	4	7	38	38	7	48	42	79	11
Fit.....	14	31	14	31	0	45	45	0	14	76	59	31
Van.....	8	37	8	37	1	44	44	1	9	81	52	38
Lat.....	18	27	18	27	0	45	45	0	18	72	63	27
Pow.....	28	17	28	17	14	31	31	14	42	48	59	31
Wri.....	36	9	36	9	21	24	24	21	57	33	60	30
Bur.....	39	6	39	6	19	26	26	19	58	32	65	25
Total.....	681	264	681	264	136	809	809	136	817	1,073	1,490	400
Positive quality first.				Negative quality first.				Grand Totals.				

TABLE XXXV

DISTRIBUTION OF JUDGMENTS. CHILDREN

Observers	Positive Quality First				Negative Quality First				Total Distribution			
	1st	2d	Pos.	Neg.	1st	2d	Pos.	Neg.	1st	2d	Pos.	Neg.
Aye.....	39	6	39	6	6	39	39	6	45	45	78	12
Bio.....	34	11	34	11	4	41	41	4	38	52	75	15
Dec.....	41	4	41	4	8	37	37	8	49	41	78	12
Bil.....	41	4	41	4	5	40	40	5	46	44	81	9
Gil.....	36	9	36	9	3	42	42	3	39	51	78	12
Col.....	39	6	39	6	3	42	42	3	42	48	81	9
Gil.....	45	0	45	0	0	45	45	0	45	45	90	0
How.....	40	5	40	5	11	34	34	11	51	39	74	16
Smi.....	6	39	6	39	3	42	42	3	9	81	48	42
Sau.....	43	2	43	2	1	44	44	1	44	46	87	3
Total.....	364	86	364	86	44	406	406	44	408	492	770	130
Positive quality first.				Negative quality first.				Grand Totals.				

and from these results the distribution with respect to positive and negative qualities might be calculated. But since in one case the positive judgments would coincide with those directed toward the first, and in the other case with those directed toward the second stimulus, the source of the totals in such a table would not be at once clear. Consequently, for the sake of clearness, the two types of distribution are given, in parallel vertical columns. The numbers in the two columns will be the same, the difference being in their arrangement.

TABLE XXXVI

DISTRIBUTION OF JUDGMENTS IN THE VARIOUS MODALITIES OF SENSATION.
TEACHERS AND COLLEGE STUDENTS

Modality or Situation	On 1st	On 2d	On Positive	On Negative
Lifted weights.....	54	72	113	13
Length of lines.....	57	69	108	18
Size of squares.....	52	74	99	27
Volumes.....	56	70	97	29
Pitch of tones.....	44	82	91	35
Shades of gray.....	34	92	89	37
Amounts of money.....	56	70	97	29
Degree of pain.....	53	73	112	14
Perfumes, affective tone..	61	65	120	6
Dates.....	60	66	85	41
Pressures.....	55	71	110	16
Intensity of sounds.....	61	65	120	6
Surfaces, texture.....	62	64	87	39
Speed of metronome.....	51	75	70	56
Bad odors.....	61	65	92	34
Total judgments.....	817	1,073	1,490	400

TABLE XXXVII

DISTRIBUTION OF JUDGMENTS IN THE VARIOUS MODALITIES. CHILDREN

Modality or Situation	On 1st	On 2d	On Positive	On Negative
Lifted weights.....	24	36	54	6
Length of lines.....	27	33	55	5
Size of squares.....	26	34	56	4
Volumes.....	27	33	53	7
Pitch of tones.....	26	34	42	18
Shades of gray.....	28	32	50	10
Amounts of money.....	26	34	56	4
Degree of pain.....	31	29	49	11
Perfumes, affective tone.	30	30	58	2
Dates.....	27	33	35	25
Pressures.....	27	33	57	3
Intensity of sounds.....	29	31	55	5
Surfaces, texture.....	26	34	50	10
Speed of metronome....	24	36	46	14
Bad odors.....	30	30	54	6
Total judgments.....	408	492	770	130

If there is no inclination to prefer the first or the second, the positive or the negative stimulus, there will be a chance distribution of the judgments with respect to the stimulus which becomes the subject of the proposition expressing the judgment. If there is a constant tendency to direct the judgment toward either the first or toward the second stimulus presented, there will be of necessity an equal number of positive and negative judgments, since both qualities occurred the same number of times in the second and first orders of presentation. If however there is instead a constant tendency to direct the judgment toward either the positive or the negative stimulus, these judgments will be for the same reason distributed between the first and second positions. What is really found is summed up in the following table.

TABLE XXXVIII
SUMMARY OF DISTRIBUTION

Observers	Positive Quality 1st		Pos. Neg.		Negative Quality 1st		Pos. Neg.		Grand Totals			
	1st	2d	1st	2d	1st	2d	1st	2d	1st	2d	Pos.	Neg.
Adults.....	681	264	681	264	136	809	809	136	817	1,073	1,490	400
Children...	364	86	364	86	44	406	406	44	408	492	770	130
Totals.....	1,045	350	1,045	350	180	1,215	1,215	180	1,225	1,565	2,260	530

The grand totals show that there is no striking preference for either the first or the second position. Such difference as is present, is about 6 per cent. more than chance relation in favor of the second stimulus presented. This balance is due chiefly to the cases in which the positive stimulus comes second, in which case there are only 180 judgments on the first as compared with 1,215 on the second. When the positive is presented first there are on the contrary 1,045 judgments directed toward the first stimulus as compared with only 350 toward the second. The direction of the judgment is not determined to any considerable degree by the mere fact of temporal position.

But examination of the tendency toward positive and negative quality shows that here there are very decided preferences and inclinations. There are a total of 2,260 positive judgments, as compared with only about 25 per cent. as many negative judgments (530). The tendency toward the positive holds no matter in what order the stimuli are presented. However, along with the pronounced inclination toward the positive quality, there is, as pointed out above, a slight preference for the second position as such. Consequently when the positive is second in order of presentation, the ratio of positively directed judgments to those negatively directed is very large (6.8 to 1). When the positive is presented first the ratio is smaller, but is still pronounced (3 to 1).

This inclination toward the positive quality is more striking in the case of the children than it is with the adults, the final ratio for the former being about 6 to 1, and for the latter 3.7 to 1. The children, that is to say, show less inclination toward the second stimulus as such and more inclination toward the positive quality as such than is the case with the adults.

The members of the group of adults show practical uniformity in this inclination. The final results for the 21 individuals show not a single exception to the general rule. Only when the positive comes first and the slight inclination toward the second stimulus favors the negative quality are any exceptions shown. Then the judgments of four adults show the reverse relation and one individual shows an impartial distribution.

A similar uniformity characterizes the group of children. In the final totals there is no exception to the general rule. When the positive is presented first a single individual with a strong inclination towards the second stimulus, affords the only exception in the table.

Tables XXXVI. and XXXVII. show the distribution of the judgments of both groups with respect to the modality or situation within which the stimuli fall. With respect to the slight preference for the second stimulus, all of the 15 groups of stimuli agree. With the adults this tendency is most pronounced with the shades of gray and the pitch of tones, and least pronounced with surfaces, sound intensities, perfumes and disagreeable odors. With the children it is most pronounced with the weights and speeds, while odors and perfumes show no difference, and pains are slightly reversed.

With respect to positive or negative direction again, all modalities and situations agree. With adults the inclination toward the positive is most striking with perfumes, sound intensities, pains, weights, and pressures, the ratio here being about 10 to 1. It is least evident with speeds, dates, and surfaces, although even here the ratio is as high as 2 to 1. In the case of the children the positive-negative ratio is highest with perfumes and pressures, and lowest with dates, pitches, and speeds.

In Table XXXIX. the various modalities have been grouped into five sections according to the degree of positive tendency shown. Thus group 1 contains the three modalities or situations which show the most pronounced inclination toward the positive quality, section 5 containing the three which show the least tendency. The figure after each modality shows the section into which that group falls. That the order of the various modalities for the two groups of observers is almost identical is shown by the fact that the modalities fall into much the same section of the total series of 15, for both

groups. Those which stand high with the adults stand high with the children also, and the positions in the scale practically coincide, so long as the same tendency is under consideration. But modalities standing high for inclination toward the second stimulus tend, of course, to fall low for inclination toward the positive quality.

TABLE XXXIX

Modality	Inclination Toward the Second Stimulus		Inclination Toward the Positive Quality	
	Adults	Children	Adults	Children
Weights.....	2	1	1	3
Lengths.....	4	4	2	2
Squares.....	2	2	3	1
Volumes.....	3	3	3	3
Pitch.....	1	1	4	5
Grays.....	1	4	4	4
Money.....	3	2	3	2
Pain.....	2	5	2	4
Perfumes.....	4	5	1	1
Dates.....	4	4	5	5
Pressures.....	3	3	2	1
Sounds.....	5	4	1	2
Surfaces.....	5	2	5	4
Speeds.....	1	1	5	5
Bad odors.....	5	5	4	3

Several interesting points are to be noticed here with reference to what the positive quality is felt to be in the different situations. With the grays it is darkness, not brightness, that is the positive quality. With dates it is recency, and still more curiously, even with the stale cheese odors, which most observers felt to be unpleasant in character, the positive quality, as indicated by the direction of the judgments, is agreeableness just as was the case with the pleasant perfumes.

Such facts as these suggest that what we have called the "positive quality" of a modality or of a judgment situation is not a permanent or characteristic property of that modality or situation throughout its whole range, but depends perhaps on the *absolute impression* received from the selections presented. This would mean, then, that if the grays, for example, which were presented in the experiment, had been lighter grays than those actually used, the observers might perhaps have received an absolute impression of brightness rather than of darkness, and that this absolute impression would modify the natural inclination of the judgments.

This form of absolute impression would, however, be somewhat different from the absolute impression which plays a rôle in the comparison of stimuli in a given experimental series. Experiments are

under way which are designed to determine whether the selection of stimuli from the extremes or middle of the scales of magnitude, intensity, brightness, affective quality, etc., reveals any change in the preferred direction or inclination of judgment, at what points the changes come, if present, and what individual differences are shown by various observers. These results will not be presented in the present connection. The purpose of the experiments here reported was simply to determine whether or not, under the conditions of a given judgment situation, definite, characteristic, and uniform tendencies of judgment expression and direction of attention are present. That such is the case, and what the character of these tendencies is, have been clearly indicated. The chief results may be summarized as follows:

Summary

1. The most striking inclination shown is a strong tendency to direct the judgment toward the stimulus described as "positive" in quality. This tendency is present with both children and adults, with all modalities and situations, regardless of the order in which the stimuli are presented. The tendency is markedly stronger with children than with adults. There are no exceptions to the general rule, among the 31 observers studied. Among the various modalities and judgment situations differences are shown, which are common to both groups of observers.

2. There is a slight tendency to favor the second stimulus presented. This inclination is not nearly so strong as the positive tendency, is weaker with children than with adults, and is consistently stronger in some modalities than others. It is strongest in those modalities in which the positive inclination is weakest.

CHAPTER VII

JUDGMENTS OF SIMILARITY AND DIFFERENCE¹

WHEN an observer is presented with two stimuli and instructed to compare them with respect to some general property such as weight, size, pitch, affective quality, intensity, etc., it is apparent that he has fairly decided preferences or inclinations with respect to the form in which his judgment is expressed. Thus comparisons of weight may proceed in terms of either heaviness or lightness, comparisons of pitch in terms of either highness or lowness, comparisons of affective quality in terms of either agreeableness or disagreeableness. But experiments show (see Chapter VI.) that judgments in terms of lightness, lowness, shortness, smallness, faintness, etc., are very infrequent so long as the observer is left to his own inclination. These categories, which may be designated as "negative," since they imply the absence of some positive factor in the stimulus or situation, seem to be, if not more artificial, at least more unaccustomed than the contrasting and grammatically opposite "positive" categories.

Conceivably these natural tendencies or inclinations or judgment habits may exert an appreciable influence on the apperception of the two stimuli, and hence on the outcome of the judgment in cases in which the differences, though objective, are small. This point has not remained untouched in the technique of the psychophysicists. As we have seen in Chapter V., Brown emphasizes the fact that, in the comparison of lifted weights, the judgment of difference depends upon the form of expression. It will be recalled that Müller and Schumann, and Müller and Martin made certain recommendations as to procedure in psychophysical experiments, as a result of related observations.

When Brown's report appeared the writer was in the midst of an investigation of judgments of a "subjective" type, such as are involved in the comparison, estimation, and measurement of such complex material as handwriting, comic situations, arguments, appeals to instincts and interests, photographs, etc. One of the problems outlined in that investigation (the results of which comprise, in part, the present monograph) is that of investigating the influence of the category or form of expression on the outcome of judgments of similarity and difference, and of other pairs of logical or grammatical

¹ Reprinted from *The Psychological Review*, September, 1913.

opposites, of analyzing the psychological relation between the two types of judgment, and of discovering the relative ease, consistency, and certainty of the various categories when the judgments are directed toward the same material, both in the case of the same observer and with groups of observers. The present chapter concerns itself with the first mentioned pair of categories,—*similarity* and *difference*.

The problem, in the writer's mind, grows at once out of the contradictory character of the few relevant references available in the literature of judgment. The following references to experimental and general studies will illustrate the point, and raise more or less definitely the question at issue.

June E. Downey, "Preliminary Study of Family Resemblance in Handwriting."

Bulletin No. 1, Dept. of Psychology, Univ. of Wyoming.

"In general a judgment of unlikeness is made with greater ease than one of likeness" (p. 49). "Toward the close of a series the judgments became judgments of dissimilarity. The records show that such a judgment is frequently made more easily than is a judgment of likeness. . . . There were subjects . . . who were more constant in their judgments of dissimilarity than in those of similarity, and who varied less from the average in the case of the latter. Some subjects . . . first selected the specimens most unlike the standard and then proceeded to find the similar hands by elimination of the unlike" (p. 20). "The judgment of unlikeness is, on the whole, an easier one to make than the judgment of likeness. There is considerable agreement among subjects as to the handwriting most unlike a given specimen" (p. 24), etc.

These statements are based on the variabilities of five successive trials by the same individuals, the instructions being "to arrange the writing specimens in the order of their likeness to a given standard" (p. 15). But if one is judging in terms of *likeness* one can not fairly speak of judgments of *unlikeness* resulting from such an experiment. It is assumed here that the category in which the judgment is expressed has no influence on the outcome of that judgment. But I shall show later that a judgment of unlikeness is not merely the reverse of a judgment of likeness, but a new kind of judgment. The "least similar" is not therefore the "most unlike."

George V. N. Dearborn, "Notes on the Discernment of Likeness and Unlikeness." *Journal of Philosophy, etc.*, February 3, 1910, p. 57.

Reports a research which "aimed to help the analysis of the mental process by which we become aware of similarity and dissimilarity . . . judgments as to the likeness and unlikeness experienced in the case of a series of visual forms. . . . The method of experimentation in detail was simply as follows: The hundred blot cards (bearing blots of ink) being placed in order ten-square on the table before the seated subject and the norm in its frame conveniently before his eyes and above the blots, he proceeded to select within fifteen minutes the ten blot-cards out of a hundred most similar in form or shape to the norm,

and to place them one side arranged carefully and deliberately in the order of their judged similarity to the norm. Meanwhile the subject reported how he apperceived the norm and what he considered its most essential form-characteristics and peculiarities. These subjective notes were recorded and the numbers of the ten blots judged most like the norm, and in their chosen order. The time required for a selection satisfactory to the subject was also recorded, and at the end of the selection the reason why each of the ten had been preferred, concisely as possible. The process in the case of judgments as to unlikeness was precisely the same, with the appropriate change in intention to keep dissimilarity instead of similarity in mind" (pp. 57-58).

Dearborn continues: "Ideal criteria (as distinguished from affective) gave more accurate results in the dissimilarity choices than in the similarity choices. This is as we should expect on logical principles. The awareness of unlikeness is an easier, if not a simpler, process apparently than that of likeness, for the change of consciousness is greater and so easier to appreciate. At any rate the sets of blots chosen as unlike the norm were much more certainly unlike it than were the 'similar' blots chosen like it" (p. 61).

There are two things to be pointed out in this connection. The first is the fact that in Dearborn's experiment the judgments of likeness and of unlikeness were directed toward totally or partially different stimuli, and hence the ease of the judgment as mere judgment is in no way indicated by his results. It may well have been that the dissimilar blots differed from the standard in more points than that number in which the similar blots resembled the same standard. In the absence of quantitative measurements of amounts of likeness or unlikeness, the relative ease of the two types of judgment can be made out only when the same material is employed in the two cases. The second point is that the assumption that the awareness of unlikeness is a simpler and easier process than the awareness of likeness seems to the writer to be completely gratuitous, until the difference has been experimentally demonstrated. The results of the present experiments indicate that the contrary is the case.

As opposed to the point of view suggested in the two articles just referred to, we find in other places frequent assertion of the more fundamental character of the judgment of resemblance, and the derived character and secondary importance of the judgment of difference. Thus Miss Macdonald, in her review of Preyer's "*Infant Mind*," says that likeness is more easily discerned than difference.

Titchener, "*A Text Book of Psychology*," p. 26, says: "We notice these differences (in human bodies) because we are obliged, in everyday life, to distinguish the persons with whom we come in contact. *But the resemblances are more fundamental than the differences.* If we have recourse to exact measurements we find that there is in every case a certain standard or type to which the individual more or less closely conforms and about which all the individuals are more or less closely grouped. And even without measurement we have evidence to the same effect; strangers see family likenesses which the members of the

family can not themselves detect, and the units in a crowd of aliens, Chinese or negroes, look bewilderingly alike.' That there may be a difference in the psychological character of the two judgments is suggested by the same writer's statement that "reports of equality or identity are less frequently based on imageless comparison than reports of difference" (p. 534).

Jevons, "Principles of Science," pp. 43 and 44, insists that similarity and difference are only two forms of expression of one and the same judgment. "In every act of intelligence we are engaged with a certain identity or difference between things or sensations compared together." "We can not, in fact, assert the existence of a difference without at the same time implying the existence of an agreement." "Agreement and difference are ever the *two sides of the same act of intellect*, and it becomes equally possible to express the same judgment in the one or the other aspect." "It is a matter of indifference in a logical point of view, whether a positive or a negative term be used to denote a given quality and the class of things possessing it." "But there are very strong reasons why we should employ all propositions in their affirmative form." "All inference proceeds by the substitution of equivalents and a proposition expressed in the form of an identity is ready to yield all its consequences in the most direct manner. . . . Difference is incapable of becoming the ground of inference; it is only the implied agreement with other differing objects which admits of deductive reasoning, and it will always be found more advantageous to employ propositions in the form which exhibits clearly the implied agreements."

Bergson, "Creative Evolution," p. 214, remarks: "Independently of all consciousness the living body itself is so constructed that it can extract from the successive situations in which it finds itself the similarities which interest it, and so respond to the stimuli by appropriate reactions." Also (pp. 44-46): "We must have managed to extract resemblances from nature which enable us to anticipate the future."

The last three references seem to agree on the proposition that psychologically, in real life, it is similarity that most interests us. If we perceive difference it is only for the sake of a search for similarity—conformity to type, interest, image, desire, etc. In handling coins the differences usually lapse in favor of the similarities, except in the case of the expert. To perceive differences requires special, sometimes professional training, and this is not necessarily because the differences are *smaller* than the agreements. They may be just as obvious, once they become interesting. We are seeking for agreements. In hunting, the resemblance of the stubble to the form of a rabbit is more striking than its many points of difference. So in diagnosing disease we are strongly interested in certain diagnostic features and accustomed to look for them, since *they* are significant in the midst of infinite diversity of other factors. Just as we are prone to "see only those instances which are favorable to the theory or belief which we already possess" (Creighton, "Logic," p. 250), so we tend to warp every perception toward the idea or image which we happen to have at the time. And just as in observing a race of men, the members of a profession, or a species of animal or plant life,

we tend always to form a conception of a type or mode from which the separate members of the group shall vary the least, so in so artificial a task as the process of judging the separate magnitudes of an experimental series we tend to conceive a central value from which the total deviations of the different magnitudes shall be the least. The clearly demonstrated "central tendency of judgment," the so-called "indifference point phenomenon" *may* be due largely to the fact that resemblances are more striking than differences, and hence all magnitudes approximating the type are assimilated towards it (see Chapter IV.; also, "*The Inaccuracy of Movement*," Ch. III., on "The Indifference Point").

The Present Experiment

The purpose of the experiment was to investigate the influence of the category or form of expression on the outcome of judgments of similarity and difference, to analyze the psychological relation between the two types of judgments, and to discover the relative ease, consistency, and certainty of the two judgments when directed toward the same material, both in the case of the same observer and with groups of observers.

The material to be judged consisted of 35 specimens of handwriting, each specimen written by a different individual, the individuals chosen at random. Each individual wrote, on a standard sized card, the words,

Department of Psychology
Barnard College
Columbia University.

One individual wrote two copies, one of which served as the standard by which the other 35 specimens were judged. The same cards and the same standard were used throughout the experiment, which covered a period of 14 months.

The chief observers, nine in number, were divided into three groups, designated by the words "similarity 1st," "difference 1st," and "mixed." Each member of the first group proceeded as follows. He was given the pack of 35 specimens, accompanied by the standard card. He was asked to arrange the cards in an order of *resemblance to the standard*, placing the most similar specimen at the top, the *next most similar* in the second place, and the *least similar* at the bottom, with the remaining cards in their appropriate intermediate positions. After completing his arrangement, for which he was allowed all the time desired, he was handed a sheet of paper and re-

quested to give an introspective account of the criteria used in passing his judgments. A week later he was again given the cards and asked to again arrange them in an order of *similarity* to the standard. After this second arrangement he was given his introspection sheet and asked to note down any modifications of criteria observed in this second trial.

After another week the same observer was given the cards and asked to arrange the specimens of handwriting in an order of *difference from* or *unlikeness to* the standard, putting at the top of his list the card *most different*, at the bottom the card *least different*, etc. A fresh introspection sheet was prepared after this arrangement, and criteria noted without reference to the previous records. After a third week a second arrangement on the basis of *unlikeness to the standard* was made, and further notes made on the introspection sheet.

The "difference 1st" group performed the experiment in the same way, except that their first two arrangements were in terms of *difference* and the last two in terms of *similarity*. In the case of the "mixed" group an arrangement on the basis of similarity was followed by an arrangement for difference, or vice versa, before the second trial for the same category of judgment.

Only one of the observers (H. L. H.) knew the purpose of the experiment at the beginning. One observer (Str.) suspected the purpose before his arrangements had all been made. Observer H. L. H. repeated the four arrangements 14 months after the first trials had been made. The intervals of one week seemed to be sufficiently long to eliminate any very decided memory effect except in the cases of the one card written in the same hand as the standard, and one other card which was strikingly different from that standard in almost every respect.

The place of each card in the various orders was recorded for each observer. The data secured from such a procedure can be examined from many points of view. In the case of each observer the two orders for *similarity* can be correlated, and the consistency of such a judgment indicated by the coefficient of correlation. The same thing may be done with the two orders for *difference*. The orders for *difference* may be inverted and the reciprocal order thus obtained correlated with the original orders for *similarity*. In the same ways may be treated the final orders for both *similarity* and *difference* secured by averaging the arrangements of the nine observers. The three groups of observers may be compared with each other in all these respects. In the case of the final orders for both

categories, the variability of the individual judgments may be computed for each card, and the categories and groups of observers again compared with respect to this *variability of judgment*. The arrangements of the various observers may be compared with the final orders secured from the group averages, and in this way the agreement of each individual with the group average (judicial capacity) determined. Comparing these measurements with the correlation between the various trials of the same observer affords a measure of the relation between *personal consistency* and *general judicial capacity*. Various other interesting and perhaps significant comparisons may be made, some of which will be later pointed out. All of these points of view will throw light on the psychological relation between the two categories of judgment, which it is the main purpose of the investigation to study.

The results of many of these comparisons and correlations are given in the following tables. In computing coefficients of correlation the formula

$$r = 1 - \frac{6(\sum d^2)}{n(n^2 - 1)}$$

has been used. The introspections of the observers, in so far as they bear on the point of the experiment, are also given.

Table XL gives the coefficients of correlation between the various arrangements of each of the nine observers, along with the average coefficients for the group. In this table S1 indicates the first trial for *similarity* and S2 the second trial. D1 and D2 indicate the two trials for *difference*. Whenever *similarity* is correlated with *difference* the reciprocal of the *difference* order (the inverted order) is used.

TABLE XL

CORRELATIONS BETWEEN THE VARIOUS ARRANGEMENTS BY THE SAME INDIVIDUALS

S, Similarity. D, Difference. The orders for difference were inverted whenever similarity was correlated with difference. The figures represent positive coefficients of correlation, by formula given in text.

Subject.	S1 with S2	D1 with D2	Average	S1 with D1	S2 with D2	Average
L.S.H., S 1st. . .	.833	.813	.823	.639	.723	.681
DeN., S 1st. . .	.781	.572	.677	.619	.655	.637
Str., S 1st. . .	.700	.811	.756	.506	.767	.636
Rich., D 1st. . .	.856	.676	.766	.654	.740	.697
Bar., D 1st. . .	.748	.586	.667	.613	.653	.633
G.E.H., D 1st. . .	.916	.727	.822	.630	.754	.692
Hart, Mixed. . .	.756	.775	.765	.572	.784	.678
Kup., Mixed. . .	.771	.894	.832	.760	.911	.835
H.L.H., Mixed. . .	.744	.677	.710	.439	.495	.467
Average.789	.726	.757	.604	.720	.662
Mean variation . .	.052	.087	.055	.065	.079	.061

Several points are at once disclosed by Table XL.

1. The correlations of the two arrangements according to *similarity* ($S1$ with $S2$) are greater than the correlations of the two arrangements for *difference* ($D1$ with $D2$). With six of the nine observers this is clearly the case. With three it is not true. Two of these three are in the mixed group, and in one of these cases there is no real difference between the two coefficients. The third exception to the rule is in the case of observer Str., who suspected the purpose of the experiment and whose introspective account states that he was disturbed by having read up on the subject. However observer H. L. H. was aware of the purpose of the experiment from the beginning, he being in fact the writer, and his coefficients show the normal relation. Apparently the mixed order of arrangements introduces factors or tendencies not present with the other two groups (see also introspections of observer Kup. under "difference"). Whether *similarity* or *difference* is judged first, five of the six observers in these two groups show considerably higher *personal consistency* when judging *similarity*. Averaging the nine observers yields a coefficient of .789 for similarity as against .726 for difference.

2. If there is no psychological difference between judgments of similarity and judgments of difference,—if, as Jevons states, "Agreement and difference are ever the two sides of the same act of intellect, and it becomes equally possible to express the same judgment in the one or the other aspect," the inverted order for *difference* should show the same correlation with a direct order for *similarity* as do two arrangements for *similarity* or two arrangements for *difference*. The fact that the coefficients for similarity are higher than those for difference suggests that the two categories of judgment are not psychologically the same. But the case is still more apparent when these reciprocal correlations are compared with the direct ones. Observe the correlations of $S1$ with the inversion of $D1$. With every observer these coefficients are smaller than those for two arrangements for similarity ($S1$ and $S2$). The average coefficient is almost 20 per cent. lower. And with seven of the nine observers these coefficients are also lower than the coefficients for two arrangements for *difference*, the average coefficient being 12 per cent. lower.

3. With every observer the coefficient for $S2$ with $D2$ is higher than for $S1$ with $D1$, the average difference being 12 per cent. That is to say, with practise and repetition the two judgments come to resemble each other, and the inverted order for *difference* to agree more closely with the direct order for *similarity*. This, we may assume, accounts for the uncertainty shown by the members of the

mixed group, with whom the two categories clashed more quickly than with the other observers, who had made two arrangements under one category before the other category was suggested. But even in these correlations of *S2* with *D2*, six observers show less agreement than with the two arrangements for *similarity*. The average is some 7 per cent. lower than the average for *S1* and *S2*, and about the same as the average for the two orders for *difference*. Averaging the direct correlations and comparing this coefficient with the average for the inverted correlations shows a superiority of 13 per cent. in favor of the former, and among the nine observers the only exception to this rule is Kup. in the mixed group, whose two averages are identical.

It seems to be clear then, that the two categories are not merely "the two sides of the same act of intellect"; that different psychological processes are involved,—processes so different that they modify the outcome of the judgment; and further, that judgments of *similarity* are made, if not more easily, at least with higher consistency than are judgments of *difference*.

Table XLI. gives the variability of the group averages for each of the four arrangements. The average deviation of the individual judgments from the average position of each card have been calculated. It seems unnecessary to give this figure for each of the 35 cards, hence the total series of 35 has been divided into 7 sections of 5 positions each and the average of the M. V.'s of each of these sections of 5 positions is given in the table. It should be noted that corresponding sections do not always contain the same cards, although this is in general true of the two orders for resemblance and the two orders for difference.

TABLE XLI

THE VARIABILITY OF THE GROUP AVERAGES FOR THE VARIOUS ARRANGEMENTS

The figures are the average M.V.'s of successive groups of five cards.

Positions	Similarity 1st Trial	Similarity 2d Trial	Difference 1st Trial	Difference 2d Trial
1 to 5 inc.....	5.18	4.46	4.70	5.40
6 to 10 inc.....	5.44	6.78	6.42	7.90
11 to 15 inc.....	6.76	6.88	7.76	7.77
16 to 20 inc.....	6.34	7.72	8.16	7.58
21 to 25 inc.....	6.34	6.96	6.78	5.82
26 to 30 inc.....	7.58	6.14	5.76	7.42
31 to 35 inc.....	5.26	4.76	4.56	4.54
Average.....	6.13	6.24	6.30	6.63
M.V.....	.71	.82	1.11	1.19

In this table then we are dealing no longer with personal consistency but with the variability of a group of nine observers. Two facts of interest are disclosed by this table. The first is that, although

the final averages of the variabilities under the four trials differ very little, such differences as are present point to *lower variability* for *similarity* judgments than for judgments of *difference*. Both the averages for *similarity* are lower than either of the averages for *difference*. There seems to be a slight tendency for the second trials to be more variable than the first, although the difference is small and not reliable. But such as it is, this difference is greater in the case of the *difference* series than in the case of the *similarity* series.

The second fact disclosed by the table is that with the arrangements for *similarity* the cards at the top of the series show smaller variability than those of the corresponding section at the bottom,—thus the first five tend to be less variable than the last five, the second less than the sixth, the third than the fifth. But with the arrangements for *difference* the reverse tends to be the case,—that is, the sections below the center of the series are less variable than the corresponding sections above the center. What this means then is this: that whether judging in terms of *similarity* or in terms of *difference*, it is on the cards which are *most like* the standard that the judgments of the various members of the group of observers agree most closely.

Summing up the results of this table we may say that the observers agree with each other more closely when judging *similarity* than when judging *difference*, and that in either case they agree more closely on the cards which are *more like* the standard than on those which are *more unlike* it.

The results of the two tables just discussed are further confirmed by those shown in Table XLII. One observer made arrangements of the cards for both *similarity* and *difference* fourteen months after the original experiment, not having examined the cards in the meantime. These arrangements have been correlated with the similar arrangements of the original experiment. The correlation between the original and the later orders for *similarity* was .69. That for the original and the later order for *difference* was .62. But the correlation between the original order for *similarity* (*difference*) and the inversion of the later order for *difference* (*similarity*) was only .36 (.62). That is to say, with an interval of over a year, personal consistency for *similarity* is somewhat higher than that for *difference*, and the difference between the one category and the inversion of the other is present and is especially striking in the case of the first two arrangements of each period.

The final group average orders for the four arrangements have been correlated, and Table XLII. presents these coefficients also. They are all four extremely high, and the differences between them are so small as to afford no suggestions.

TABLE XLII

MISCELLANEOUS CORRELATIONS OF ARRANGEMENTS

Correlations of final group average orders:	
1st order for similarity, with second trial.....	93
1st order for difference, with second trial.....	95
1st order for similarity with reciprocal of 1st order for difference.....	93
2d order for similarity with reciprocal of 2d order for difference.....	91
Subject H.L.H., correlations of trials 14 months apart:	
1st resemblance with resemblance 14 months later.....	69
1st difference with difference 14 months later.....	62
1st order for resemblance with reciprocal of order for difference secured 14 months later.....	36
1st order for difference with reciprocal of order for resemblance secured 14 months later.....	62

In the following pages are given the introspections secured from the nine chief observers whose results have been recorded, and also introspections from several others who were asked to make but one arrangement, some for *similarity* and others for *difference*. A discussion of the significance of these introspections will follow them.

*Introspections**Resemblance:*

DeN.—The principal thing upon which my judgment was based was the general slant of the writing,—that is the sample was in a hand slanting from right to left and the ones slanting in the same general direction looked more like it than the vertical or backward. Another thing was the formation of the capitals, especially of the letters P and C. Another factor was the space between the letters,—whether the word was all connected or whether it was broken.

Kup.—At first the actual combination of various types of hand writing, *e. g.*, slant, round, backhand, as evidenced in the type given as a model appealed to me and I was inclined to sort the cards according to this “combination type.” Soon, however, the elements of character, of the personality in back of that type copy claimed my attention and this criterion established itself in my mind as a standard by which to judge the others. I characterized the type copy as having elements of rapidity, definiteness, free movement and no-waste-of-time. It seemed that of a decided, quick thinking person. According to such characteristics I tried to arrange the cards given.

Hrt.—The first resemblance I thought of was that of slope, then the question as to whether the joinings between the letters were sharp or curved. Then I compared the relative height and depth of the letters, above and below the lines. Then I noticed endings of words, whether they ended abruptly or with a flourish. Methods of crossing t's and dotting i's were noticed and also methods of finishing y's and g's. The apparent ease of the writing always struck me,—whether it seemed to swing along easily or to be stiff and cramped. The size of the letters received little attention on the whole.

Rich.—My introspections are just about the same as when I arranged the cards for difference instead of resemblance, except that instead of looking to see how the cards differed in general appearance, placing, slant, color, etc., I looked for similarity in these respects.

Bar.—I was influenced primarily by regularity or irregularity of lines in the writing. If the whole seemed to be made up of lines going in all directions I was inclined to classify it as like the standard. If the whole presented an orderly appearance I did not consider it like the standard. I was influenced also by the width and prominence of the pen line—choosing, first those that were darker and heavier, like the standard. Sometimes I found myself comparing only the one word “psychology” on the various cards,—then when I tried to see them all at once—the factor of regularity or irregularity was the strongest. Slant had some influence, but the judgment was much a matter of *general impression*, without any special factor so prominent. The ideas were mainly impressionistic,—I was guided more by a feeling of like or unlike than I was by any specific comparisons.

Str.—I first grouped the cards according to the position on them of the three lines of writing, then according to *uniformity*, regardless of the style or legibility, and finally, when the cards were very poor, according to legibility.

L.S.H.—I based my judgments of similarity to the standard on the shape of the letters and the slant of the writing.

H.L.H.—Began in terms of slant and judged on basis of slant, roundness of letters and general appearance of the card, until about two thirds of the way down. Then the slants were all reversed, the judgments seemed more difficult and the criterion was shifted to letter formation,—angles, tails of y’s, capitals, becoming more important. On turning back to the start, after the first arrangement, these later factors asserted themselves, and I rearranged the first few cards, paying more attention to the smaller details than I had done before.

Cas.—The general character of the writing, as a whole, was the main basis for the arrangement. By that I mean the general size, boldness or fussiness and regularity. Next in importance was slant, and then the formation of the various letters.

Wund.—I judged first by the general character of the writing, then by the slant of the letters, the distance the letters were apart, and their general roundness. As I reviewed my first arrangement I made several changes according to the resemblance of the final letters of the different words, noting whether they turned up or down. I also watched for the ways in which the t’s were crossed.

Lyo.—Personally I think I more or less unconsciously considered several factors, such as shade of ink, position on card, legibility, script, and size,—I said “this or that card is *like* the standard” without forming the reason in words.

And.—First on the type of handwriting,—an extremely masculine type,—then on the slant of the letters and lastly on their form.

Hod.—I based my judgment chiefly on the general appearance and direction of the writing, whether it was slanting, upright or backhand. I took into consideration also the size of the writing, the spacing of the letters and the form of the letters themselves.

Wd.—In the first place I tried to pick out handwriting with the same general slant and carelessness and arrangement. Then I noticed the capitals and then of the endings of the words, the spacing and the size of the letters, although these latter I did not use very much. The general features seemed more important to me than the smaller details.

Difference:

DeN.—I paid more attention to the formation of independent letters than when I arranged the cards for resemblance. Used slant until about one third of

the way through then had to rely on minor details, and the task became harder.

Kup.—This arrangement was constantly harder than the previous one, because of my inclination to arrange as I had done last time when the order was that of resemblance. When instinctively I felt the great difference of a card I very often remembered that I had not placed it so low in the order for resemblance. I labored between two impulses,—one to be true to my previous judgments and the other to act honestly according to my present light. I think I succeeded in following the latter. I noticed as I had not done before, to so great an extent, the great resemblance of groups of cards. Very often they seemed to have been written by the same person, but with the intention to disguise his handwriting. In such cases I noticed the *details* of the penmanship and made my decision rest with such little points as the separation of letters in a word, the crossing of a t or the last stroke of the y. . . . Throughout the relation of resemblance was in the background of consciousness. I felt that it was involuntarily more a criterion than the standard of "difference." The problem seemed far more puzzling this time than last.

Hrt.—In ranking according to dissimilarity I did not think first of slope, as in the arrangement for resemblance, but rather of differences in endings of letters like g, y, etc., and in beginnings of words after capitals.

Rich.—I first looked at the general type of writing, *i. e.*, the slant, the size of the letters and the blackness of the ink. After this more general survey I thought sometimes of the similarity of the formation of the letters and the capitals, but this was necessary only when the general survey did not show striking enough differences.

Bar.—First the general appearance of the writing in its suggestion of the character of the writer. The pattern seemed to express a type of individuality entirely different from that expressed in the card which I placed on top. This is a question of general impression. For cards more nearly alike I think the strongest point was in the regularity or irregularity of the letters. Some seemed to be regular according to some definite system, others, like the sample, seemed to be more or less hit-or-miss style. Another feature was the width of the pen line. Next came the question of slant, although this was not a very strong factor. The formation of the individual letters was also of small import, but the final letters of each word influenced me somewhat, also the capitals. The question of motor imagery seemed to be a determining factor,—I seemed unconsciously to wonder how differently one should go about it to write the various cards, and to think of the hand movements necessary to the writing. This was a very strong factor in judging those that were particularly dissimilar.

Str.—Judged by general conception of smoothness rather than by actual comparison of standard. This may have been due to the fact that I had just read Dearborn's article on "The Discernment of Likeness and Unlikeness." Found the judgment harder than that of similarity and laid more stress on details which went to make up general smoothness. Distasteful job, goes counter to normal mode of doing things. Tended for a while to think of similarity. Do not feel sure of my judgments.

L.S.H.—Felt less decided than when making judgments of resemblance. Judgments vaguer. Felt as though about to come down stairs backwards, and thus a little uncertain of progress. Judgments based on slope, shape and size of the letters with some tendency to consider the "maturity" of the writing.

H.L.H.—Began in terms of general slope and "rapidity." Felt rather in the air and soon found the criterion inadequate. Then adopted size for a while,

then formation of separate letters, tendency to flourish, and way of ending y's, g's, and d's. In the last part the tendency to think in terms of resemblance was strong, because the cards resembled each other in slant of the letters. Had to use finer and finer details.

Wood.—I judged first on the form of the letters and the way in which they were made, then on the general direction,—vertical, slant or backhand. Then the position of the words on the card, and finally such details as the crossing of the t's, the ending of the y's and the way the e's were made.

Gold.—My judgments were chiefly based on differences in slant, size, and heaviness. My first judgments were made by examining the writing, as a whole, comparing one card with another. Later I studied the individual words and letters, comparing their shape, roundness or sharpness, whether connected or not, method of crossing t's, etc.

Read.—In deciding the differences in handwriting the first consideration was the general appearance. So long as the cards of decided vertical writing held out I went by that. I then noticed the differences in the formation of the letters and particularly the first and last letters of a line. Of course, to some extent, the general effect was still of influence.

Grand.—I first observed the general character of the writing. The standard seemed to me to be freely flowing, accustomed and not particularly careful. I began selecting those cards which were most carefully and apparently most slowly written, and those which seemed to have been written with some difficulty. As the most striking cards were eliminated the process became more difficult and I paid more attention to the formation of individual letters.

Plum.—The factors considered were general neatness, angles and slant, size of the writing, arrangement of the lines on the cards, and the form of special letters, such as the d and the C.

Two things are indicated with considerable clearness by these introspective records. The first is the greater ease and naturalness which is felt to characterize the judgments of similarity. This is best revealed in the introspections made during arrangements for difference. Thus Kup. reports: "This arrangement (difference) was constantly harder than the previous one (similarity). . . . The problem seemed more puzzling this time." Str. records: "Found the judgment harder than that of similarity. . . . Distasteful job, goes counter to the normal mode of doing things. Tended for a while to think of similarity. Do not feel sure of my judgments." Similarly L. S. H. remarks: "Felt less decided than when making judgments of resemblance. Judgments vaguer. Felt as though about to come down stairs backwards, and thus a little uncertain of progress." H. L. H. reports: "Felt rather in the air, . . . found the criteria inadequate . . . tendency to think in terms of resemblance was strong."

The second fact is suggested by such statements as often occur when judging difference,—"I paid more attention to the formation of independent letters than when I arranged the cards for resemblance" (DeN.). Or, "I noticed the details of penmanship and made my decision rest with such little points as the separation of letters . . . , the

crossing of a t or the last stroke of a y (Kup.). Also "I did not think first of slope, as in the arrangement for resemblance, but rather of differences in endings of letters like g, y, etc., and in beginnings of words after capitals" (Hrt.). "Began in terms of general slope and rapidity . . . and soon found the criteria inadequate" (H. L. H.). "I judged first on the form of the letters and the way in which they were made" (Wood). The judgment of difference, that is to say, is largely or often based on the comparison of fine points and minor details.

The introspections for similarity, on the other hand, abound to a much greater degree in references to "slope," "general slant," "character," "personality," "regularity," "uniformity regardless of the style or legibility," "general impression," "carelessness," etc.—all of these factors of a large, general, loosely defined and "impressionistic" character. These differences in criteria tend to assert themselves without regard to the order in which the arrangements were made.

A possible objection at this point might be that the differences in the two arrangements were perhaps due to the fact that the two arrangements began with different cards (the similar end of the series in one case and the unlike end in the other), rather than to a real influence of the form of the judgment. A test of this would be afforded by observers who should arrange the cards in terms of *similarity* (beginning with the most similar) and also in terms of *difference* (beginning with the least different instead of with the most different). When such an experiment was tried with three observers, all three showed clearly that, in the attempt to reason out what might be meant by "least different," the two categories were at once brought explicitly together in the consciousness of the observer. Since logically the "most similar" is the "least different," the arrangement then proceeded in terms of similarity, even when the instructions were in terms of difference.

The apparent objection is not a real one. The observer has all the cards before him. Whatever cards are judged to be "least similar," he may leave till the latter part of the series, if he chooses, when judging similarity. When judging difference, whatever cards he judges to be most different may be at once selected. The whole matter is in the observer's own hands. And the significant thing is that the cards which are left to the end of series, when judging similarity, are not precisely the ones selected for the earlier part of the series when judging difference.

Furthermore, if the result were only a consequence of inverting the series, the two orders for difference should correlate as closely

as, and show no greater variability than, the two orders for similarity. Neither of these conditions is realized. The difference is then not merely the result of inverted arrangements.

Summary

1. The personal consistency correlation of two arrangements on the basis of *similarity* is greater than that of two arrangements for *difference*, unless, by performance in the "mixed order," or by some other circumstance, both categories are brought explicitly together in the consciousness of the observer.

2. Both the correlation of two orders for similarity and of two orders for difference are higher than the correlation of an order for similarity with the reciprocal of an order for difference.

3. With repetition, adaptation and familiarity with the material the two categories tend to approximate each other and the direct order to agree more closely with the indirect order.

4. The variability among a group of observers is less for similarity than for difference.

5. Whether the judgment is expressed in terms of similarity or in terms of difference it is on the cards which are most like the standard that the group agrees most closely.

6. When arrangements are made 14 months apart, the same relations are disclosed,—personal consistency for judgments of similarity is greater than that for judgments of difference, and the discrepancy between the direct order and the indirect order secured by inverting the arrangement under the opposite category is noticeable.

7. Introspection suggests the greater "ease" and "naturalness" and "confidence" of the judgments of similarity.

8. Introspection also shows a different distribution of criteria in the two categories. Judgments of similarity tend to be based on grosser and more general criteria, such as character, slope, ease, rapidity, etc.; the judgment tends to be "impressionistic." In judging difference more attention is paid to the finer details of form, size, arrangement, and separation of letters.

9. Judgments of similarity and of difference are not merely two forms of expression of one and the same intellectual act. Judgments within each type or category involve each its own peculiar psychological processes and criteria. The "most similar" is not, by virtue of that fact, the "least different," nor is the "least similar" identical with the "most different." Of the two categories, similarity seems to be the most fundamental, natural, easy, and self-consistent, whether a single individual or a group of observers is concerned.

10. In these respects judgments of similarity and of difference behave in the same way as do judgments of other logically opposite qualities (such as preference and dislike, intelligence and stupidity) which involve, in the beginning of such an experiment, psychological processes and criteria which are not identical, but which move to a common plane as the experiment proceeds or is repeated (see Chapter VIII.).

CHAPTER VIII

THE INFLUENCE OF FORM AND CATEGORY ON THE OUTCOME OF JUDGMENT¹

As we have seen in the preceding chapter, judgments of similarity and of difference are not merely the two sides of one and the same act of intellect, but involve each its own peculiar psychological processes and criteria, and the category or the form in which the judgment is expressed, the attribute toward which it is directed, makes a considerable and measurable difference in the outcome of that judgment. The present study reports an investigation, from a similar point of view, of certain other judgments commonly passed in daily life.

Is a judgment of stupidity the exact reverse of a judgment of intelligence? Is a judgment of preference the exact reverse of a judgment of dislike? In other words, do we use the same standard in judging characteristics designated by logical opposites, ranking all specimens according to the degrees by which they deviate positively or negatively from that standard? When we arrange specimens of handwriting in an order of merit with respect to *resemblance* to a given standard hand we use somewhat different criteria from those employed when the specimens are arranged according to their *difference* from the standard. May it be also true that judgments of intelligence or of preference are based on different sets of criteria from those of judgments of stupidity or aversion? Do we like a person for certain qualities and dislike those who possess the exact antithesis of these qualities, or are our dislikes and preferences based on different sets of qualities? To discover which of these possibilities has the greater degree of probability is the main purpose of this study.

The material consisted of 25 photographs of actresses. The photographs were similar in shape, size, finish, and mount, differing only with respect to the individual photographed and the pose assumed. In selecting the photographs care was taken to avoid those of well-known actresses, in order that past judgments might not influence the results of the experiment. These pictures were ranked in an order of merit, by 10 observers, with respect to *preference*, *dislike*, *intelligence*, and *stupidity*. As the purpose was to discover the

¹ By Margaret Hart Strong and H. L. Hollingworth. Reprinted from *Jour. Phil., Psych., and Sci. Methods*, September 12, 1912.

effect of the direction or category of judgment, special emphasis was laid on each category in the written instructions with which each of the observers was provided. These instructions were as follows:

Preference

Arrange the photographs in an order of merit, placing at the top the face you *like the most*, placing second the *face you like next best*, and so on, until the face you *like the least* is at the bottom of the series.

Dislike

Arrange the photographs in an order of demerit, placing at the top the face you *dislike the most*, placing second the one you dislike next intensely, and so on, until the one you *dislike the least* is at the bottom.

Intelligence

Arrange the photographs in an order of merit with respect to the *intelligence* of the face, putting at the top the *most intelligent*, next to it the next in intelligence, and so on, with the *least intelligent* face at the bottom of the series.

Stupidity

Arrange the photographs in an order with respect to the *stupidity* of the face, putting the *most stupid* at the top, next to it the next stupid, and so on, until the *least stupid* looking face is at the bottom of the series.

Five of the observers made the arrangements in the following order:

- 1st week, ranked for preference and intelligence.
- 2d week, ranked for preference and intelligence.
- 3d week, ranked for dislike and stupidity.
- 4th week, ranked for dislike and stupidity.

The remaining five ranked for dislike and stupidity in the first two weeks, and for preference and intelligence in the last two weeks. This precaution was taken in order to minimize the influence of practise on the results of the group averages. In every case at least a week intervened between one judgment and the next. There was no clear evidence of decided memory effect except in the case of the extremes of the series. After the fourth arrangement the observers were asked to write out a statement of the criteria used in judging each trait. The observers were all students of Barnard College, juniors or seniors taking their second or third year's work in psychology.

In making the correlations to be discussed later, the formula

$$r = 1 - \frac{6(\sum d^2)}{d(d^2 - 1)}$$

was used. The correlations were worked out between each observ-

er's two trials (I. and II.), and between each observer's average judgment (*a*) with the group judgment (*A*), for each of the four traits. These results are given in Table XLIII.

TABLE XLIII

THESE COEFFICIENTS OF CORRELATION ARE ALL POSITIVE

Observer	Ell.	Car.	Ste.	Hal.	DeN.	Str.	Bro.	Bar.	Val.	Cas.	Av.	M.V.
Correlations of I. and II.:												
Preference.....	55	73	87	91	68	74	88	92	84	96	80.8	10.6
Dislike.....	57	89	86	98	87	73	84	70	86	60	79.0	11.0
Intelligence.....	71	84	90	92	78	74	86	77	91	83	82.6	6.0
Stupidity.....	77	85	89	87	83	72	73	65	82	86	79.9	6.5
Correlations of <i>a</i> with <i>A</i> :												
Preference.....	51	57	58	23	56	55	44	45	54	58	50.1	7.7
Dislike.....	50	59	64	31	43	27	57	48	63	48	49.0	9.6
Intelligence.....	32	29	32	48	43	41	32	59	26	30	37.2	8.4
Stupidity.....	54	57	55	52	62	46	62	36	42	36	50.2	8.2

Table XLIV. gives the correlations between each order and the reciprocal of its supposed opposite (by the reciprocal is meant the inverted order, so that what was originally the bottom of the series becomes the top). If categories logically opposite are also psychologically the two sides of the same act of intellect, then the correlation between preference and the reciprocal of dislike should be equal to the average of the personal consistency coefficients for preference and for stupidity. That is to say, the inverted order for dislike should coincide with the direct order for preference, and should correlate as closely with this direct order as would two trials for preference with each other. The same relation should be expected to hold between intelligence and stupidity. On the other hand, if the processes differ from each other psychologically, it would seem that the correlation between preference and the reciprocal of dislike (both standards or categories being involved) should be less than the corre-

TABLE XLIV

Observer	Ell.	Car.	Ste.	Hal.	DeN.	Str.	Bro.	Bar.	Val.	Cas.	Average
Correlations of:											
1. Pref. and the recip. of disl.	60	89	93	94	90	57	86	78	89	83	81.9
2. Av. of pref. I. and II., and disl. I. and II.....	56	81	86.5	94.5	77.5	73.5	86	81	85	78	79.9
3. Int. and the recip. of stup.	85	79	93	90	94	74	73	87	86	96	85.7
4. Av. of int. I. and II., and stup. I. and II.....	74	84.5	89.5	89.5	80.5	73	78.5	71	86.5	84.5	81.2

lations of two trials for preference or of two trials for dislike. The same, again, should hold for intelligence and stupidity.

At first glance, as the results are presented in this table, the situation does not seem to be similar to that found in the study of judgments of *similarity* and *difference*. In 6 of the 10 cases the correlation between preference and the reciprocal of dislike is greater than the average correlations of similar arrangements, and in two of the remaining cases there is no difference between the two. The average shows a small per cent. in favor of the former.

In the case of intelligence and stupidity, 7 of the 10 observers have higher correlation between the judgment of intelligence and the reciprocal of stupidity than the average correlation of similar arrangements, and the average shows superiority in this direction of 4.5 per cent.

It is apparent then that if these character judgments really have the same psychological differences as those found between judgments of similarity and difference, some factor is present in this experiment which obscures the difference.

Table XLV. indicates that this factor is practise, adaptation, or familiarity with the material, and that before these factors operate genuine psychological differences are disclosed. In this table the trials are not averaged as in Table XLIV., but the first order for preference is correlated with the reciprocal of the first order for dislike. and the second order for preference with the reciprocal of the second order for dislike. In a similar way are handled the arrangements according to intelligence and stupidity. Each of these indirect correlations is then compared with the average of the direct correlations,—that is, with the average of preference with preference, and dislike with dislike. This also is done in the case of intelligence and stupidity.

In both cases the results are clear. The correlation of the first of the positive quality with the reciprocal of the first of the negative quality is less than the average correlation of positive and negative qualities with themselves. In the case of preference and dislike there is no exception to this rule, and the average difference amounts to over 13 per cent. In the case of intelligence and stupidity 3 of the observers are exceptions, but the other 7 show the difference clearly; a difference which averages, for the 10 observers, over 5 per cent. Averaging the two types of judgment, in the lower part of the table, there is no exception to the rule, and the average superiority amounts to over 9 per cent.

The influence of practise, adaptation, and familiarity with the

material is shown by comparing the third row of coefficients in each group of Table XLV. with the second row of the same section. In these third rows the correlation of the second direct arrangements with the second of the reciprocal arrangements is seen to move up, in each case, and very clearly in the average, to the correlation of two direct arrangements for a given trait. In fact the coefficients

TABLE XLV

Observer	Ell.	Car.	Ste.	Hal.	DeN.	Str.	Bro.	Bar.	Val.	Cas.	Average
Av. pref. (I. and II.) and disl.											
(I. and II.).....	56	81	87	95	78	74	86	81	85	78	79.9
Pref. I. and recip. of disl. I....	22	81	83	91	66	43	77	56	80	67	66.6
Pref. II. and recip. of disl. II...	59	80	90	95	92	55	79	86	82	90	80.8
Av. int. (I. and II.) and stup.											
(I. and II.).....	74	85	90	90	81	73	79	71	87	85	81.2
Int. I. and recip. of stup. I....	72	78	88	88	87	53	52	73	77	92	76.0
Int. II. and recip. of disl. II...	83	78	88	90	91	69	86	84	83	87	83.9
Av. pos. and neg. (I. and II.).	65	82	88	92	79	73	82	76	86	81	80.5
Pos. I. and recip. of neg. I....	47	80	86	90	77	48	65	65	79	80	71.3
Pos. II. and recip. of neg. II..	71	79	89	93	92	62	83	85	83	89	82.3

are usually a little higher. Very evidently, then, in the beginning of the experiment, before the two categories have been brought together in the consciousness of the observer in any explicit way, the judgment of a negative quality is not the exact antithesis of that of a positive quality. A judgment of dislike, that is to say, is not merely the reverse aspect of a judgment of preference, but a new kind of judgment, with perhaps different criteria, and certainly with a different outcome. The same must be said of judgments of intelligence and stupidity. The form of expression, the direction or category of the judgment, has a measurable influence on the outcome of that judgment. But as the experiment proceeds and the two categories are both explicitly brought to the consciousness of the observer, and after practise, adaptation and familiarity with the material have played their part, the difference between the two categories tends to fall away, and the form or direction of the judgment no longer influences its outcome.

This tendency is the same as that remarked in the study of the judgments of similarity and difference in the case of handwriting, where it is found that with practise and repetition the two judgments come to resemble each other, and the inverted order for *difference* to agree more closely with the direct order for *similarity*.

This tendency is further shown by the figures in Table XLVI., in

which the correlation of the first two trials of a given observer is compared with the correlation of his last two trials, regardless of the category of judgment concerned. With a single exception the latter coefficient is always higher than the former, the average of the ten observers showing a superiority of 7 per cent.

TABLE XLVI

Observer	Ell.	Car.	Ste.	Hal.	DeN.	Str.	Bro.	Bar.	Val.	Cas.	Average
First two trials.....	63	79	89	92	73	73	79	68	84	73	77.0
Last two trials.....	67	87	88	93	85	74	87	85	88	90	84.2

TABLE XLVII

PERSONAL CONSISTENCY COMPARED WITH GENERAL JUDICIAL CAPACITY

Observer	Ell.	Car.	Ste.	Hal.	DeN.	Str.	Bro.	Bar.	Val.	Cas.	Average
Average correlations of I. with II.	65	83	88	92	79	73	83	76	86	81	80.6
Average correlations of <i>a</i> with <i>A</i>	47	51	52	39	51	42	49	47	46	43	46.6

TABLE XLVIII

Ratio of Best to Poorest	Preference	Intelligence	Dislike	Stupidity	Average
Correlation of I. and II.....	96:55	92:71	98:57	89:65	1.51:1.00
Correlation of <i>a</i> with <i>A</i>	58:23	59:26	64:27	62:36	2.15:1.00
Average.....					1.83:1.00

TABLE XLIX

Correlations of					
I. and II.:	Av.	M.V.		Av.	M.V.
Preference.....	80.8	10.6	Subjective judgments...	78.9	10.8
Intelligence.....	82.6	6.0	Objective judgments....	81.3	6.2
Dislike.....	79.0	11.0	Positive judgments.....	81.7	8.3
Stupidity.....	79.9	6.5	Negative judgments....	79.4	8.8
<i>a</i> with <i>A</i> :					
Preference.....	50.1	7.7	Subjective judgments...	49.5	8.6
Intelligence.....	37.2	8.4	Objective judgments....	43.7	8.3
Dislike.....	49.0	9.6	Positive judgments.....	43.7	7.9
Stupidity.....	50.2	8.2	Negative judgments....	49.6	8.9

The introspection was of little value, consisting for the most part of mere generalization. But where specific criteria were given the presence of the two standards was apparent. For example, Observer Hal.—“I like eyes looking straight at me. I don’t like head or eyes to have unnatural pose, because it looks affected. I can’t abide frowsy hair. I like smiling eyes and mouth and a high forehead.” Here the first two criteria do seem to be opposed—eyes looking straight at one are not usually eyes in an unnatural pose. But other criteria show the two standards. The observer “can’t abide” frowsy hair, but she does not specifically admire smooth coiffures. She likes high foreheads, but expresses no positive dislike for low ones.

Some incidental points brought out in the results are worth noting. In Table XLVII. the personal consistency of each observer is compared with her correlation with the group average. The coefficient (.06) shows that there is absolutely no correlation between the two. This seems to indicate an absence of general judicial capacity.

In Table XLVIII. the ratio of best to poorest is given, and the familiar ratio of about 2:1 found (see Chapter X.).

Table XLIX. seems to show that the more subjective judgments of preference and dislike are more variable and uncertain than the more objective ones of intelligence and stupidity. The coefficients are slightly lower on the average and the mean variations are larger. This is true whether personal consistency or judicial capacity is concerned. The coefficients for the negative judgments of dislike and stupidity also show a higher variability than do those of the positive judgments of preference and intelligence.

Summary

1. Judgments which are grammatically opposite (as preference and dislike, intelligence and stupidity) involve, in the beginning of the experiment, psychological processes and criteria which are not identical. The form, direction, or category of the judgment exerts a measurable difference on its outcome.

2. As the experiment proceeds the processes and criteria move to a common plane and the two types of judgment resemble each other more closely. This movement to a common plane is apparently the result of repetition, adaptation, and familiarity with the material, and of the fact that the two categories, hitherto implicitly distinct from each other, are now brought explicitly together in the consciousness of the observer.

3. The result of practise and familiarity with the material is to increase the personal consistency of the observer's judgments.

4. Introspection suggests different criteria for judgments which are grammatically or logically only two sides of the same intellectual act.

5. There is seen to be no correlation between personal consistency and agreement with the group average.

6. The ratio of best to poorest, in both these respects, is the familiar one of about 2:1.

7. Subjective judgments (of preference and dislike) are more variable and uncertain than the more objective judgments (of intelligence and stupidity).

8. The coefficients of "negative" judgments (dislike and stupidity) are more variable than those of the "positive" judgments (preference and intelligence).

CHAPTER IX

THE PERCEPTUAL BASIS FOR JUDGMENTS OF EXTENT¹

IN 1887, in the course of experiments on the extent of movement, Loeb² was led to the supposition that the judgment of extent is based on the perception of the duration of the movement. Since then Kramer and Moskiewicz,³ in 1901, and Jaensch,⁴ in 1905, have felt that their experimental results led to the same conclusion. Woodworth,⁵ in 1903, discredits the hypothesis. His chief objections are: (1) Duration may be varied without entirely destroying the approximate equality of the extents; (2) extent can be judged better than time; (3) compensatory constant errors with higher speed are insufficient; (4) if we judged by duration alone, speed distinctions would be reduced to a matter of visual space or perception of force.

In June, 1909, the writer published, along with other matter,⁶ the result of a long series of experiments on the relation between the judgments of extent and duration in the case of rectilinear arm movements. His conclusion there was that "the experimental facts point to separate processes of judgment for the two magnitudes, extent and duration. The four methods of separate accuracy tests, confusion, correlation, and correction failed to justify the assumption that the perception of any one characteristic of a movement is more primitive or fundamental than that of any other. The judgment of extent seems to be based on a system of signs which have been learned to mean extent directly. The same seems to be true of both duration and velocity."⁷

In the July (1909) number of the *American Journal of Psychology*, Leuba⁸ reported experiments, on the results of which he arrives at conclusions quite opposed to those quoted in the preceding paragraph. "The comparison of the length of arm movements is made through the comparison of the duration of one or several of the sen-

¹ Reprinted from *The Journal of Philosophy, Psychology, and Scientific Methods*, November 11, 1909.

² *Pflüger's Archiv*, 41, p. 124, 1887.

³ *Zeitschrift für Psychologie*, 25, pp. 101-125, 1901.

⁴ *Ibid.*, 41, pp. 257-279, 1905.

⁵ "Le Mouvement," Chap. IV.

⁶ "The Inaccuracy of Movement," *ARCHIVES OF PSYCHOLOGY*, No. 13, 1909.

⁷ *Ibid.*, pp. 85-86.

⁸ *American Journal of Psychology*, July 1909, p. 374.

sations arising from the movement and of a particular value of the joint sensation called here the rate value."

In the face of such conflicting opinion the writer desires to present in abbreviated form the results of his experiments and to give certain additional reasons in support of his earlier conclusions.⁹ From 600 to 800 experiments were performed on each of four subjects, by the method of average error, on extents ranging from 150 to 650 mm. and on corresponding durations ranging from 1 to 3.5 seconds. By using a piece of apparatus already described elsewhere,¹⁰ all the movements, while they remained active, were free

TABLE SHOWING RELATION BETWEEN ERRORS OF EXTENT AND ERRORS OF DURATION

						<i>Deliberate</i>					
EXTENT						DURATION					
Obs.	Trials	Per Cent. C.E.	Per Cent. V.E.	Right Guesses	r	Trials	Per Cent. C.E.	Per Cent. V.E.	Right Guesses	r	
W.	450	6±2.0	13±0.6	59	.22	375	5±1.3	11±0.7	46	.31	
H.	450	19±1.7	12±0.6	54	.56	375	16±2.0	12±0.9	52	.54	
Bt.	287	24±3.8	18±1.5	64	.79	264	20±3.5	16±1.2	61	.67	
L.	375	7±0.8	7±0.6	60	.54						
Averages		14±2.1	12.5±0.8	59	.53		13.7±2.3	13±0.9	53	.51	
<i>Incidental</i>											
W.	375	8±1.7	13±0.8	49		450	10±1.8	20±0.9	53		
H.	375	9±1.3	12±0.6	56		450	8±0.9	12±0.6	58		
Bt.	264	15±2.2	15±1.2	65		287	17±2.8	20±1.3	63		
L.						375	5±1.5	13±0.9	56		
Averages		10.7±1.7	13.3±0.9	57			10±1.7	16.3±0.9	56		

from the *illusion of impact* which has vitiated so much of the work on movement. The apparatus gave simultaneous graphic records of the extent, duration, speed, and energy of every movement performed. For further details of the experiment and for a more complete presentation of most of the data used in the present article the reader must be referred to the writer's earlier monograph. The preceding table gives the C.E. and V.E. for the extents and their corresponding durations, when the observer tries to reproduce (1) the extent and (2) the duration of his first movement. In still other columns may be found the per cent. of right guesses when the observer guessed, after each trial, as to the probable direction of his error, and the coefficient of correlation between agreement of extents and agreement of durations calculated by the method of unlike signs.

⁹ Leuba's article was probably in the hands of the printer when "The Inaccuracy of Movement" appeared.

¹⁰ "Inaccuracy of Movement," Chap. I.

On the basis of these figures the writer draws the following conclusions.

1. The durations of extents intended to be equal have greater V.E. (16.3 per cent.) than the extents themselves (12.5 per cent.). There must be, then, some basis for the judgment of extent other than the perception of duration.

2. The C.E. seems to be bound up with the process of attention, the magnitude deliberately reproduced [extent (14 per cent.) or time (13.7 per cent.)] being greater than that of the magnitude incidentally reproduced [time (10 per cent.) or extent (10.7 per cent.)]. This evident separation between the magnitude attended to and that incidentally executed argues for separate processes of judgment for the two magnitudes, extent and duration.

3. If the perception of duration were the basis of the judgment of extent, incidentally reproduced durations should show as close correspondence as durations deliberately reproduced. This is not the case.

4. Extents agree as closely when the observers are reproducing duration (V.E. 13.3 per cent.) as when they are attending to the extent (V.E. 12.5 per cent.), but durations incidentally executed do not correspond as closely (V.E. 16.3 per cent.) as in deliberate experiments on reproduction of duration (V.E. 13 per cent.). That is to say, if either judgment is to be considered the more primitive and fundamental it should be the judgment of extent rather than that of duration.

5. The coefficients of correlation between deliberate extents and incidental durations (+.53) on the one hand, and between deliberate durations and incidental extents (+.51) on the other, are positive. But all that this shows is the presence of positive correlation between extent and duration, no matter which factor is being attended to. There is as much evidence for the dependence of duration judgments on the perception of extent as for the converse.

6. If the observer is required to guess as to the probable direction of his error in the case of each attempt to reproduce either extent or duration, (a) the guesses in both cases correspond more closely to the actual errors of the extents (59 per cent., 57 per cent.) than to the differences between the durations (57.5 per cent., 53 per cent.); (b) the proportion of right guesses in experiments on extent (59 per cent.) is greater than that in experiments on duration (53 per cent.). These facts are unfavorable to the hypothesis that it is the perception of duration on which the judgment of extent is based.

Leuba's chief argument is based on the proposition that the dura-

tions of movements judged shorter, equal, or longer than a standard fall out shorter, equal, or longer as compared with the duration of the standard. Unfortunately, neither the variability nor the reliability of the average is given, nor is the number of cases, from which a reader might compute the reliability himself. But even if the correspondence were found to be *complete* such statistical correspondence would throw no light whatever on the nature of the process of discrimination involved in the comparison of the two lengths. If accurate measurements had been kept of the depth of the wrinkles in the loose glove which covered the arm of the observer there would have been found the same positive correlation—when the extents were judged shorter the wrinkles would have been found to be relatively shallow, and they would have been equal or deeper according as the judgment happened to be “same” or “longer.”

It is a case in which denying one member of the disjunction disproves a conclusion which is not proved by the affirmation of the other member. In other words, even though the relations of the durations do coincide with the form of the judgment, this duration agreement may still be simply an incidental fact, on a par with the depth of the wrinkles in the observer's sleeve. With the rather constant speed characteristic of all observers in such experiments a greater extent must occupy a longer duration, an equal extent an equal duration, etc. To show that the durations *do not* agree as closely as the extents, as the writer has already done, invalidates the one conclusion, while to prove that they agreed equally well would have no bearing whatever on the question of the perceptual basis of the judgment of comparison.

The movements reported in Leuba's article were made in different parts of the arm's total swing, under different degrees of contraction, tension, joint position, etc. The only common factor was the time element. Now even to prove that under these unusual conditions the duration of movements is used as the basis for the comparison of their extent does not prove that this is what happens in other cases. But to show that even here the durations disagree more than the extents disproves the hypothesis completely.

With Leuba's assertion of the existence of a special set of signs which serve as criteria for judgments of speed, the writer heartily agrees, but he is convinced that along with this assertion should also go the recognition of the independent character of judgments of extent and duration.

CHAPTER X

SOME CHARACTERISTICS OF JUDGMENTS OF EVALUATION

AMONG the most common judgments passed in daily life are those which express preferences or aversions, similarities or differences, convictions or doubts, successes or failures, and other "general impressions" or value "estimates." These expressions possess all the characteristics of judgments, but are often said to be "subjective," in the sense that it is impossible or difficult to measure their truthfulness or accuracy by the application of a standardized test. In many cases no "objective" (generally accepted or conventionalized) measure exists, and the only method of test is by observing the internal consistency of an individual's judgments on different occasions, by comparing the individual's judgments with the consensus of opinion of a large experimental group of observers, or by some other statistical criterion. In such cases there is, strictly speaking, no measurement of truth or accuracy, but rather of the consistency, certainty, frequency, or correlation of different judgments.

The dependence of these judgments of general impression on individual differences gives them a particular psychological interest. Esthetic and ethical judgments belong to this group, as do also many verdicts in the fields of philosophy, politics, manners, justice, and most of the decisions of business, pedagogy, and religion. In spite of the practical importance of this type of judgments, experimental psychology has until recently occupied itself with only the more trivial of them. The evaluation of simple esthetic material,—the elements of design, color preferences, tonal harmony, and the various attributes of elementary sensory experiences have been studied in detail. But there have been few attempts to investigate experimentally the characteristics, conditions, and behavior of judgments of such qualities as eminence, interest, belief, persuasion, character, the comic, literary merit, etc.

Studies conducted by the "methods of expression" may be disregarded in this connection, since these methods are expressly directed toward the facts and character of the organic reaction rather than toward the characteristics of the accompanying process of judgment. Of the "methods of impression" various forms have been developed, such as the "method of paired comparisons," the "serial method," "order of merit method," etc. In the hands of different investigators

these various names have not always meant precisely the same procedure, but the general features of the methods are well recognized. Perhaps the most conspicuous have been the methods of "paired comparisons" and "order of merit." Of these two the latter is by far the more promising and Miss Barrett (1) has recently demonstrated its superiority from the points of view of simplicity, expedition, and reliability and significance of results. The present paper considers some of the characteristics of such judgments of evaluation as those for which the "order of merit" method has been used in the past.¹

The beginnings of the method may be seen in some of the simple experiments of Fechner, Mantegazza, and Galton. The method was first given definite formulation by Cattell in a study of brightness intensities (2) and particularly in his statistical studies of eminent men and women (3-7). The method has since been used and further developed by many of Cattell's students, including Summer (21). Norsworthy (17), Wells (24, 25), Thorndike (22, 23), Strong (18, 19), Kuper (16), Barrett (1), and the writer (11-14). Downey (8) and Yerkes (26) have also employed the method, and Thorndike (23) has further proposed the transmutation of results secured by this method into a surface of distribution for the purpose of deriving quantitative statements of amounts of difference.

In most of these studies the method has been used chiefly as an instrument in the investigation of some specific problem, such as family resemblance, interests of children, value of advertisements, measurement of school progress, distribution of eminence, etc. But when the various studies are considered as a group there arise a number of interesting problems concerning the judgments themselves. Certain of these problems will here be taken up in turn, with a brief consideration of the data at present available for their solution and interpretation. In many cases the conclusions can be but tentative, and in several cases the problems themselves may ultimately prove to be but "straw problems," suggested by a chance coincidence of accidental or insignificant results. In spite of these facts it seems worth while to present the problems in a more or less definite way, in order that future results may be explicitly referred to them.

Many of these problems were first suggested directly or indirectly in the two very original papers of Wells. The general principle of the method may be given in the words of this author. "Professor Cattell calls attention to the fact that, if one endeavors to arrange

¹ For full bibliography of these studies see end of chapter.

and rearrange in serial order a number of given objects, the positions successively given them will vary somewhat as they would vary if the arrangements had been made one each by different observers. If we undertook to arrange ten times a series of grays in order of brightness, we should no more get the same order each time than we should get identical orders from ten different subjects. Nor would our own orders vary approximately the same amount from the average; sometimes we should be better, sometimes worse, judges, just as among our ten subjects some would be more discriminative, some less. The judgments of the same individual at different times are theoretically quite comparable to those of different individuals regardless of the factor of times" (25—1).

A fuller description of the method and illustrations of some of its useful practical applications are to be found in the writer's "Principles of Appeal and Response" (14). A further modification, which may be designated the *group method* as contrasted with the *strict order method* has been employed by the writer, and possesses several advantages which justify its further development. The following account of this modification is taken from a previous paper (11).

"Instead of arranging the material in strict order of merit the observer placed them in ten piles, according to their 'degree of funniness.' In the first pile were placed the superior jokes, in the tenth the poorest ones, while the intermediate piles represented gradation of merit from best to poorest. No instructions were given as to the amount of difference represented by these successive piles, nor as to the number of cards to be placed in each.

Ten observers took part in the experiment, all of whom were women, students in the Barnard laboratory, with one and a half year's work in psychology. When the average position of each card for the ten observers was calculated, the 39 jokes could be arranged in a strict order of merit according to their respective averages. The advantages of this group method are several.

* It is much quicker than the strict method, less fatiguing and monotonous to the observer, yet correlates closely with results from the same observers by the strict order method. Further, the method gives opportunity to observe any changes in value of the group as a whole. Thus by multiplying the number of cards in a given group (say 7) by the position of that group (say number 9) and adding these products for all ten groups a figure is obtained which gives some measure of the total value of the series for a given individual or group. Now if the cards are arranged a second, third, fourth, etc.,

time by the same observers, these sums will indicate the change in total value of the series during the successive trials. This figure is of course not in any sense an absolute measure. It is conditioned by shifts in the individual's standard of value, by his personal variability of judgment, by the variation in standard from individual to individual, and by the fact that no card can be thrown higher than the first nor lower than the last pile. Nevertheless it affords an interesting and suggestive index of the total series behavior which the strict order method can not yield. It will be shown later that the M.V. (mean variation) in such experiments bears a constant ratio to the number of places into which the objects are to be sorted, so that the relative variability is the same here as in the strict method.

There may be, in the *group method*, a certain tendency to arrange stimuli according to qualitative or type resemblance, which might to a degree disturb the judgment of merit,—a tendency, that is, to put all puns in the same pile, etc. But there is no evidence in the results that such an inclination has in any way operated. Moreover the tendency is just as strong, in the *strict order method*, to put qualitatively similar stimuli in the same region of the scale. Thus Wells found that in arrangements of picture postals according to preference there was a tendency to place near each other cards bearing similar scenes, color schemes, etc. It is conceivable that, even in arranging individuals with respect to scientific eminence, contiguity in space or similarity of field or method may operate as a more or less significant associative factor in determining relative position. But since these factors also help determine the individual's actual judgment of merit, they need not be supposed to warp that judgment in any undesirable way.

In the present experiment each of the ten observers arranged the cards five successive times, the trials being a week apart. This plan thus gave data for investigating the variability of the group, of the individual, of the total value of the series, and of the behavior of each card under the influence of repetition. Both Wells and Downey have shown that a week is ample time for the elimination of any great disturbance through the memory factor in the successive trials."

Problems

First Problem. Variability of Different Parts of the Series. (Repeated arrangements and arrangements by different individuals.) —If all the items are arranged at each trial the variability of each item from its average position may be determined. When this is

done the variability is usually found to be smaller at the extremes of the series than in the central section, in such material as has been employed. The variabilities increase fairly regularly as the central region of the series is approached. The following records (Table L.) illustrate this tendency. The figures are taken from various studies in which different material and observers were used, and include series of various lengths. The results are not always given for each item, but usually for sections of neighboring items, the sections being determined sometimes by tabular convenience, and in other cases by the way in which the results were originally expressed.

Wells remarks, on this finding in the case of repeated arrangements by the same observer: "We find, as we should anticipate, that the M.V. increases toward the middle position and decreases toward the ends. The amount of this increase varies considerably and constitutes a not uninteresting point of individual difference. In subject *A* the middle M.V.'s are nearly three times those at the start, in *D* they are barely half again as much. Individual difference in reliability of judgment seems therefore to be greater in the middle than at the ends. This is what we should expect, for the judgments are more difficult in the middle and we naturally vary more from each other in our judgment of difficult things than in our judgment of easy ones" (25—525).

But the problem can not be so easily disposed of. In the first place the decrease of variability toward the ends is in part a purely methodological consequence,—items at extreme top and bottom of the series can be displaced in successive arrangements or by different observers, in only one direction, viz., toward the middle. Even those somewhat further in from the extreme ends can suffer large displacements in one direction only, but at the middle of the series there is double opportunity for large displacement. To be sure the maximum possible displacement is greater in the case of the extremes, since a given card may be displaced the full length of the series, but this situation probably seldom occurs,—would, in fact, occur only in arrangements on the basis of chance. The individual differences pointed out by Wells are then in all probability only differences in variability in general, rather than in specific "amount of increase" from one part of the series to the other.

The problem as it now stands is to determine to what extent the increase of variability toward the center is only a methodological result of this *end error*, and how far it possesses any further significance. One can not by any means assume *a priori* that in a given series the middle region will be one of greater difficulty. In fact one

TABLE L

VARIABILITY IN DIFFERENT PARTS OF THE SERIES

Study	Av. M.V. of Sections, from Top to Bottom									
	1	2	3	4	5	6	7	8	9	10
H.L.H. Jokes Funniness 39 items 10 Obs.	1.89	2.04	1.85	2.20	2.07	2.58	2.14	1.81		
H.L.H. Appeals Persuasiveness 50 items 50 Obs.	9.76				11.44					9.80
H.L.H. Portraits Intelligence 20 items 10 Obs.	1.41	2.85	3.86	3.68	3.60	3.01	2.90	3.03	2.06	2.16
H.L.H. Portraits Courage 20 items 10 Obs.	2.80	3.27	3.38	5.08	5.50	3.34	3.34	2.67	3.29	3.12
Wells Post Cards Preference 50 items 5 Obs.	8.7	8.3	11.6	10.5	12.2	12.9	10.0	10.8	11.8	8.5
Wells Authors Style 10 items 10 Obs.	.25	.30	.36	.39	.40	.39	.34	.31	.33	.26
Strong Advertisenemts Persuasiveness 10 items 30 Obs.	1.9	1.4	2.0	2.5	2.8	2.8	1.6	2.0	2.3	1.5
Downey Handwriting Resemblance 37 items 10 Obs.	4.72	6.58	6.50	7.43	7.03	5.94	4.48	3.62		

might expect the difficulty to increase regularly toward one end of the series, unless the material were deliberately chosen so as to afford items on both sides of the zero-point of the quality being judged. In the case of the post cards this may well have been the case, and the series may have included positively pleasing and positively displeasing as well as indifferent items. In Wells's study of the series of weights with constant difference ratios between adjacent items, the variabilities increased from the top to the bottom of the series. The same thing was true of Cattell's lists of eminent men, though here there was no lower limit to the series.

Test experiments might be made in which the presence of a zero-region could be introspectively reported upon, with different materials and varying series lengths. Only by such experiments may the rôle of the *end error* be separated from other suspected influences. The figure of variability has been used as a measure of the amount of difference between the items judged, and whenever this is done it is important to be sure that other conditions are not influencing the size of the coefficients. The table just given indicates that the tendency toward increased variability in the central region is present with varied kinds of material, regardless of the manner in which it is chosen. It will be shown later that the average M.V. of these experiments with judgments of "general impression" tends to be about one fifth of the total number of places in the series. This would mean that the *end error* might of itself affect the upper and lower quarters of the total series, which perhaps sufficiently explains the tendency to increase toward the center.

Second Problem. Certainty of Individual Likes and Dislikes.—

Disregarding the middle of the series the variabilities of the two extreme sections may be compared, since both these sections are equally affected by the *end error*. Two cases must be distinguished here: (1) The consistency or certainty of repeated arrangements by a single observer; (2) the agreement or disagreement of various individuals of a group. On the first point the following data are available (Table LI.). In this table the first section is to be compared with the last, the second with the penultimate, and the third with the antepenultimate section. It will be observed that the same individual is, on the average, more certain (has smaller M.V.) in the case of the lower sections of the series than in the case of the upper ones. With respect to his data Wells remarks: "Another point of significance is that the M.V.'s are always less at the disliked end than at the preferred end, although there is no intrinsic reason why they should be better grounded in memory. This might be in great part due to a

TABLE LI
CERTAINTY OF INDIVIDUAL LIKES AND DISLIKES

Section	H. L. H. Judgments of the Comic. M.V.	Wells Preference for Post Cards. M.V.	Downey Resemblance of Handwriting. M.V.
First.....	.84	2.6	2.69
Second.....	1.39	4.7	3.05
Third.....	1.64	5.4	3.90
Antepenult.....	1.64	5.4	2.92
Penultimate.....	1.37	4.4	2.74
Last.....	.78	1.8	1.45

generally unesthetic series of cards, but it is perhaps generally true that we are surer of our antipathies than of our preferences" (25—525). But Downey finds the same relation shown in general by judgments of resemblance, and remarks: "Toward the close of a series the judgments became judgments of dissimilarity. The records show that such a judgment is frequently made more easily than is a judgment of likeness" (8—20). The writer, in the study of judgments of the comic, finds the same tendency for the lower end of the series to show smaller variability.

Here again then is a problem. In these studies of repeated arrangements the lower end of the series shows the smaller variability. This is hardly to be explained by Wells's suggestion of the greater certainty of our antipathies, unless one can be fairly supposed to entertain feelings of aversion toward "unlikeness" when judging handwriting, and toward lack of humor in an intended comic situation. It should be pointed out that the relation is by no means a unanimous one with individual observers. Only half of Wells's observers show it to any striking degree, though all but one of the five show it when the highest five items are compared with the lowest five. In my own results the relation of the averages is largely due to four of the observers, the other six showing exactly the opposite result. One of Downey's experiments failed to show the tendency with any certainty, and the repeated arrangements of weights in Wells's study showed an increasing variability from top to bottom of the series. It is quite probable that there is no genuine problem here at all and that the results given are merely dependent on the character of the material in the particular cases. It is perhaps easier to find material that is distinctly *not beautiful*, *not comic*, or *not similar*, than to find material of the extreme opposite qualities.

Third Problem. Group Variabilities in Likes and Dislikes.—With respect to the likes and dislikes of the members of a group of

observers several studies are available. I will present first a discussion of this point as it appeared in the previous paper on "Judgments of the Comic."

"*Likes and Dislikes.*—If the cards be arranged in a final order of merit for each trial and the M.V.'s of the best cards compared with those of the poorest, that is, if the M.V.'s of the top and bottom of the series be compared, the members of the group are found to agree more closely at the top than at the bottom. Table LII. gives the M.V. for the first and last ten places in each of the five trials. Inspection shows two facts. First, that the M.V. for the top groups, taken either by 5's or 10's, is less than for the lower. Thus the average M.V. for places 1-10 is 2.03 compared with 2.22 for places 30-39. The M.V. of places 1-5 is 1.97 compared with 2.09 for places 34-39.

TABLE LII
Av. M.V.'s, 10 OBSERVERS, 5 TRIALS

Pos.	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
1	1.48	1.20	0.90	1.66	1.12
2	1.40	3.04	2.98	2.12	2.22
3	2.84	1.56	1.72	2.44	1.80
4	2.20	3.06	2.10	1.66	1.84
5	1.80	2.32	1.86	1.62	2.40
6	2.52	2.40	2.56	2.10	1.49
7	1.88	2.08	2.70	2.40	1.84
8	2.04	1.56	1.52	2.21	2.00
9	2.08	1.68	1.60	2.83	2.20
10	2.40	1.88	2.32	2.08	2.68
—					
—					
—					
30	2.60	2.76	1.43	2.80	2.40
31	3.20	2.12	1.80	2.40	2.52
32	2.08	3.04	3.18	2.80	1.96
33	2.50	2.44	2.10	2.30	1.63
34	2.08	2.12	2.24	1.60	2.17
35	1.98	2.40	2.20	1.90	1.84
36	2.94	2.20	1.68	2.40	2.38
37	2.00	1.70	3.16	1.38	1.50
38	2.36	1.88	1.80	2.50	1.56
39	2.72	1.82	1.78	1.96	1.80

Second, this difference becomes smaller with each repetition, the differences between the M.V.'s of 1-5 and 34-39 being successively .46, .23, .21, .13, .05, and between the M.V.'s of 1-10 and 30-39, being .39, .24, .17, .10, .01. Generalizing we may say that in the beginning individuals agree more closely on the good than on the

poor, but that with successive repetitions this difference disappears (see Table LIII.).

TABLE LIII
AVERAGES FROM TABLE LII

	1	2	3	4	5	Average
Av. 1-5.....	1.94	2.23	1.91	1.90	1.87	1.97
Av. 34-39.....	2.40	2.00	2.12	2.03	1.82	2.09
Difference.....	+ .46	- .23	+ .21	+ .13	- .05	
Av. 1-10.....	2.06	2.08	1.97	2.11	1.96	2.03
Av. 30-39.....	2.45	2.32	2.14	2.21	1.97	2.22
Difference.....	+ .39	+ .24	+ .17	+ .10	+ .01	

This first relation seems to be a usual one in judgments of this subjective character,—of preference, beauty, persuasiveness, etc. Thus in Wells's study of picture postals, although the author does not call attention to the fact, the figures yield the following result. For places 1-5 and 45-50, the M.V.'s are much alike, being respectively 8.7 and 8.5. For places 1-10 the M.V. is 8.5 while for 40-50 it is 10.2. For 1-15 it is 9.5 as against 10.3 for places 35-50, etc.

Various investigators find that for repeated trials by the same individual the reverse situation holds, the same individual being more consistent at the bottom of the scale than at the top, and the suggestion has been made that this may mean that we are more certain of our dislikes than of our preferences. Giving the present relation a somewhat analogous interpretation, it may mean that although a single individual may be more certain of his antipathies, a group of individuals will resemble each other more in their preferences than in their aversions.

Or the relation may mean simply that we attend to things possessing positive quality, that here where the expression of the judgment is in terms of preference we attend more strongly to the end in which our preferences really lie. But that this is not true for all individuals will be later pointed out. Dearborn finds judgments of unlikeness easier to make than judgments of similarity, and Downey finds some evidence for the same relation, although the average of her results confirms the statement of Wells. But the judgment of preference is qualitatively different from the judgment of resemblance, the one being based on feeling-tone, the other on more restricted perceptual factors.

Another possible interpretation of the data is that the differences between the superior cards, at the top of the scale, are greater than those of the mediocre at the bottom. This was clearly shown by Cattell to be the case in judgments of scientific achievement. Thus

"The figures show that the average differences² between the chemists who are in the first tenth are about eight times as great as between the chemists toward the middle of the list and about twelve times as great as between the chemists toward the bottom of the list." But there are at least three reasons for believing that there is considerable change in attitude when the same observer turns from arranging men according to merit to arranging simple stimuli according to affective tone. The difference lies in the fact that part way down the scale, in the latter case, the expression of judgment changes from terms of decreasing preference into terms of increasing positive dislike, whereas probably few scientists who would get into a total group would be rated as positively bad, the judgment being expressed rather in terms of more or less merit. Arrangements of scientific merit resemble the scale of sensation intensities, varying always in terms of degree, while arrangements of preference resemble the gradation of feelings from the positive pole through a region of indifference to a decided negative pole.

In the second place the suggestion that the smaller variability in the upper ranges depends on objective differences in the stimuli is contradicted by the fact that in the successive arrangements by the same individual four of the ten observers were more consistent in the lower range than in the upper, and this would hardly be expected if the differences between the cards in this lower range were actually smaller than in the upper. Furthermore if something like Weber's law holds for judgments of affective tone as well as for sensation intensity, differences in the upper range would have to be greater in order to yield equal variability, and considerably greater if the variability is still smaller. The whole question of this closer group agreement in the upper ranges seems to merit further investigation and especially, the tendency of the differences to become uniformly smaller in successive trials."

The following results, from the preceding chapter on judgments of similarity and difference in the case of handwriting, show the same tendency. Both when judging similarity and when judging difference the nine observers agree more closely on the upper sections of the series, the material being the same in both cases.

The following table gives the average results of two studies by Wells, the one of "literary qualities," the other of "similarity of two colors." The judgments of literary qualities show the common tendency, but the judgments of color similarities show just the reverse.

² Measured inversely by the size of the probable errors and directly by the difference in grade.

TABLE LIV

35 SPECIMENS OF HANDWRITING. 9 OBSERVERS

Trait: Resemblance to a Given Standard Specimen

Section	Judging Similarity	Judging Difference
1st 5 items,—Av. M. V.....	4.82	4.55
2d 5	6.11	6.59
3d 5	6.84	6.30
4th 5	7.03	7.87
5th 5	6.65	7.77
6th 5	6.86	7.16
7th 5	5.01	5.05

TABLE LV

	10 Authors with Respect to Given Literary Qualities. Av. M.V. of 10 Observers	28 Pairs of Colors. Average M.V. of 10 Observers
1st sec. of series.....	.25	2.1
2d sec.....	.30	2.6
	—	—
	—	—
Penultimate sec.....	.33	2.4
Last sec.....	.26	0.7

Individual and class differences in such a tendency might well be expected. In a later study by the writer, in which 50 appeals to specific instincts and interests were rated according to their persuasiveness, an apparently genuine case of such difference is afforded (12). The following table (Table LVI gives the average

TABLE LVI

Average M.V.'s of	Best 10	Middle 10	Poorest 10
20 women, 1st trial.....	10.10	11.18	10.07
20 women, 2d trial.....	9.76	11.93	9.59
10 women.....	9.37	10.58	8.77
Av. of women.....	9.74	11.23	9.47
20 men.....	9.84	12.96	10.79
Grand average.....	9.76	11.44	9.80

M.V.'s of the highest, lowest, and middle sections of 10 appeals for several groups of observers. The point of interest in these records is the question of closeness of agreement at the top of the list, among the preferences, as compared with that at the bottom of the series, among the dislikes. The evidence here is suggestive. Women seem to agree more closely on their dislikes (M.V. 9.4) than on their preferences (M.V. 9.7), but the difference is not large. It is probably reliable and genuine, however, since the relation holds in all three experiments with women. The men, on the other hand,

agree more closely on their preferences (M.V. 9.8 as against 10.8 for dislikes) and the difference is considerable. The averages of men and women show no difference whatever. There seems to be a sex difference here, which, expressed in general terms, would be, that men resemble each other more closely in their preferences while women are more alike with respect to their aversions. This fact throws some light on the further finding that there is low correlation between the magnitude of the M.V's for the particular cards when the variabilities of the women's judgments are compared with those of the judgments passed by the men.

It is difficult to determine how far this question of group variability at the extremes is merely a function of the material and how far it is due to more essential psychological factors. Such cases as the sex difference just described are obviously not due to the nature of the material, which was the same in both cases. There is further evidence which tends to confirm the suggestion of this sex difference as men and women are now constituted. Thus Strong (18—79) finds that "When women are given an equal opportunity with men to rate appeals (advertisements) they are able to classify their dislikes as well as their preferences, which the men do not. . . . Women have more and greater dislikes than men and are surer of them." Similar evidence is found in Kuper's study of the preferences of boys and girls from 6.5 to 16.5 years of age. "Another sex difference noted was the number of positive dislikes expressed by each sex. The girls gave 161 dislikes as against the boys' 65. Boys seemed to entertain relative indifference toward the appeals at the bottom of the list" (16).

These results, if further verified, would lead to the generalization that men are homogeneous, that is, tend to resemble each other more closely, in the case of their preferences,—appeals which are positive and strong; women, on the contrary tending to be alike with respect to their dislikes,—appeals which are weak or negative. Whether this difference bears in the direction of selection and difference in experience or training, or merely toward the temporary motives which operate in reacting toward such experiments, the results do not show. The fact that women have definite and mutual aversions, with fewer common preferences, while men have fewer determinate dislikes but definite and mutual preferences, is, if true, an interesting statistical discovery, and one which may be found to have numerous implications. Whether it be interpreted to mean a fundamental and inherent sex difference or merely a difference which reflects our present social organization (which is doubtless an

adequate explanation of all the facts) has nothing to do with the present usefulness of the facts themselves. Moreover the suggested further verification must be found before the existence of the difference can be asserted with even mild assurance.

Fourth Problem. Personal Consistency and Judicial Capacity.—This problem was first raised by Wells (25—529) who remarks, in discussing the esthetic judgments of his subjects, “A somewhat significant comparison is afforded between the variability of the (5) subjects from the average of the ten, and their variation from their own judgments (in repeated arrangements). Those who vary least from their own judgments also vary least from the judgments of others. . . . The observations are too few to do more than suggest a general principle, but their interpretation is a rather interesting one. The critic who best knows his own mind would seem the best criterion of the judgments of others.” In the case of the judgments of amount of resemblance between colors “the peculiar correspondence between the amount of variation from one’s own judgment and from the judgment of others appears” also.

In order to test further the truth of this generalization I have made several experiments in which the variability of the individual (*personal consistency*, as shown by the correlation of two trials by the same individual on different occasions) is correlated with his degree of agreement with the group average (*judicial capacity* or *representative character*). The resulting coefficient of correlation will thus indicate the degree to which high personal consistency implies the representative character of the judgments. The various coefficients from the different experiments are given in the following table.

TABLE LVII

PERSONAL CONSISTENCY AND JUDICIAL CAPACITY

Judgment Situation and Observers	<i>r</i>
Appeals, relative persuasiveness, 20 women29
Jokes, relative funniness, 10 women	-.49
Faces, various characteristics, 10 women06
Handwriting, resemblance, 9 observers47
Handwriting, difference, 9 observers	-.07
Syllables, agreeableness, 10 women15
Portraits, various characteristics, 10 women11
Wells, postal cards, 5 observers70
Wells, color differences, 7 observers30
Downey, handwriting resemblance, 1st specimen70
Downey, handwriting resemblance, 2d specimen	-.40
Downey, handwriting resemblance, 3d specimen40
Average	+ .19

In my own experiments, with 10 to 20 observers, the correlations are practically zero (Av. .07). I have computed, from the data given by Wells and Downey, similar coefficients from their small groups of observers, (usually 5) and these are also included in the table. Four of the five are positive and large, the other being negative, and the average being .34. The average of the 12 different studies is .19. The only large negative correlation among my own figures is in the case of the judgments of comic situations. It may well be that this single negative coefficient is due to the peculiar nature of the material. The process of adaptation gives to the comic situation a changing rather than a static value. The judgments of the group of observers in this experiment indicate that some of the jokes change greatly in value with successive repetitions. One class, the "objective comic" as I have called them (naive jokes and calamity jokes in which the predicament of the victim is self-induced) rise in the relative scale. Another class fall just as rapidly,—the "subjective comic" (sharp retort, pun, play on words, caricature, occupation joke, etc.). A third class (mixed in character) approximate their original position, in the later arrangements, and constitute about one half of the total series. This gives a waxing, a waning, and a static group.

This means that if a given individual's judgments are to be an index of the opinion of the group his evaluation of the waxing and waning items must vary correspondingly, thus giving him a low personal consistency coefficient. In so far as the individual's consecutive arrangements remain uniform, to just that extent does he fall short of being representative of his group. It is clear from these facts that in all such determinations the *stability* of the material must be in some way ascertained before the results can be safely interpreted.

Fifth Problem. Personal Consistency in Different Situations.—It would be interesting to know whether an individual who has a high personal consistency coefficient in one situation shows the same characteristic when a totally different sort of material is judged. In Table LVIII. such coefficients are given for 10 observers in two different situations,—judgments of the comic and judgments of persuasiveness of appeals. The correlation by relative position between the two columns (1 and 2 of the table) is —.30. The cases are few and the P.E. large, but in so far as the data are reliable they indicate no likelihood that an individual who judges the one sort of material consistently will judge with relatively equal consistency in the other situation. The peculiar nature of the material in these two cases gives

this conclusion merely suggestive value, and further experiments are needed.

Sixth Problem. Judicial Capacity in Different Situations (General Judicial Capacity).—The table just described contains also, for these 10 observers, their degree of correlation with the average of their group in the two experiments (columns 3 and 4 of the table). The correlation between the two columns is .22. This figure again is subject to a large P.E. In so far as it is reliable it indicates a certain degree of general judicial capacity, the individual who is the best representative of his group in the one case being somewhat more likely than any other individual to be the best representative of his group in the other situation.

TABLE LVIII
GENERAL JUDICIAL CAPACITY

Observer	Personal Consistency		Correlation with Average	
	Appeals (r)	Comic (M.V.)	Appeals	Comic
Ell.....	.55	.88	.24	.32
Mah.....	.13	1.65	.36	.55
Mor.....	.71	1.30	.13	.54
Den.....	.78	1.86	.52	.66
Ger.....	.81	.95	.66	.70
Mas.....	.87	1.43	.36	.60
Pra.....	.74	1.35	.62	.28
Bis.....	.73	.87	.43	.30
Sch.....	—	.87	—	.43
Hrt.....	.80	.92	.55	.48
$r = -.30$		$r = +.22$		

In another experiment, the results of which are not given in the table, a given group of individuals judged, on the one occasion the legibility of handwriting, and on another occasion their degree of belief in each of a series of propositions. The correlation between representative character in the two cases is just zero ($-.01$), showing consequently the non-existence of general judicial capacity in this experiment.

Wells found, in his statistical study of literary merit, that the observer who was the best judge (most nearly representative of the group) in the case of "general merit" was not at all necessarily the best judge of the author's possession of the various specific qualities. In a group of 20 observers "the worst judge of general literary merit, according to his divergences, is the third best judge of charm, the best judge of clearness, and the thirteenth best of euphony. The best judge of general merit is the fifth best of charm, the fourteenth of

clearness, and the seventeenth of euphony. . . . We can hardly draw inferences as to the general capacity for sound judgment as measured by the soundness of judgment for any particular class of objects . . . the fact that one has a good judgment for psychologists tells us very little about the value of his opinion in other fields. . . . To demonstrate the very existence of an abstract power of judgment is ultimately synonymous with the problem of free will" (24—30).

Cattell found, in the case of the judgments, by ten psychologists, of the eminence of fifty living psychologists, that "the second best judge of the first ten psychologists is the worst of the second, the fifth of the third, the eighth of the fourth, and the sixth of the fifth" (24—30). On the whole then, there is no evidence, in the available material, of the existence of such a thing as general judicial capacity.

Seventh Problem. Relation of Variability to Series Length.—Another striking relation brought out by the comparison of various order of merit arrangements of stimuli on the basis of such affective factors as preference, beauty, persuasiveness, funniness, etc., is the constancy of the ratio of the average M.V. for the series as a whole to the number of possible positions in the range. If by M.V. we designate this average variability and by P the total number of positions in the scale then $M.V./P$ is, with various kinds of material, with different groups of observers, and with a widely ranging value for P , usually .20, and with high reliability. The following table exhibits this relation in such material as the writer has at hand.

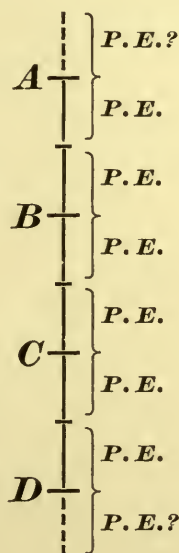
TABLE LIX

	Material	Trait	Observer	P	M.V.	M.V./ P
1.	4 advertisements	Persuasiveness	10 men	4	.8	.200
2.	5 advertisements	Persuasiveness	10 men	5	.98	.196
3.	39 jokes	Funniness	10 women	10	2.2	.220
4.	10 advertisements (av. of 4 sets)	Persuasiveness	10 women	10	2.3	.230
5.	10 advertisements (av. of 3 sets)	Persuasiveness	20 mixed	10	2.5	.250
6.	20 advertisements (av. of 2 sets)	Persuasiveness	50 mixed	20	4.3	.215
7.	20 photographs	Various traits	10 women	20	3.6	.180
8.	39 jokes	Funniness	10 women	39	8.03	.205
9.	50 appeals	Strength	20 women	50	10.5	.201
10.	50 picture postals (Wells)	Beauty.	10 mixed	50	10.7	.201

That is to say, the M.V. is always about one fifth of the total number of possible places, or the P.E. (probable error) assuming a normal distribution, about .168 or about one sixth of the range. The evidence seems to the writer too strong to permit of explanation in terms of mere coincidence. Of course if the material had been the same throughout, the only variable being the number of places into

CHARACTERISTICS OF JUDGMENTS OF EVALUATION

which it was sorted, this is just what we might expect, for the relative P.E. would remain constant, the absolute P.E. depending on the fineness of the grades of distinction. But we have here ten distinct sets of material, judged in terms of a considerable range of traits, by widely differing groups of observers, both as to sex, training, interest, and number. The only constant factor is that the judgment is always based on the affective reaction to the stimulus. And we find that in every case the probable error is approximately one sixth of the range. (It would probably be slightly larger if it were not for the fact that the *end error* tends to reduce the variability of the extreme upper and lower positions.) Assuming that the M.V.'s were equal in all parts of the range (and they do not vary greatly), and allowing a P.E. in both directions from both the upper and lower



extremes, the total range would then be divided into four sections, each separated from its neighbor by the respective P. E.'s, somewhat as follows. This would mean that, so far as the average judgment of the group of observers is concerned, there are only four distinct grades of difference or merit in the material, only four shades of distinction on which the group would, in the long run, agree, these grades corresponding to the sections lying about A, B, C, and D as central tendencies.

This situation is curiously analogous to that disclosed in judgments of the same observer, where practise shows that about four or

five distinctions of certainty, clearness, etc., are all that can be comfortably and accurately made. The same thing that holds for the variability of the individual holds for the variability of the group. And the fact that the law holds for such different kinds of material and traits argues an interesting resemblance between the judgments involved in such affective discriminations.

The size of this ratio $M.V./P$ would become smaller as the material came to be selected so as to disclose more pronounced or more objectively measurable differences. Thus in judgments of resemblance of penmanship, which are supposedly more directly perceptual and objectively verifiable in kind, Downey finds $M.V.$'s which, if arranged as below, according to the range of possible positions, would yield an $M.V./P$ value of about .163, or a probable error of about .130, meaning that while there are only about four clearly marked grades of beauty, funniness, persuasiveness, etc., there are about five clearly marked degrees of resemblance.

TABLE LX
VARIABILITY OF JUDGMENTS OF SIMILARITY (DOWNEY)

P	$M.V.$	$M.V./P$
20	3.31	.165
34	5.33	.157
37	6.22	.168

Average $M.V./P = .163$

It is probable that this ratio ($M.V./P$) can be used as a reliable index of the objective character of judgments and with greater accuracy than the crude $M.V.$ employed by Wells. Using this ratio the objectivity of his three classes of judgments would be, in increasing order,—preference .201, weights .141, colors .086, showing that the judgments of weight order were more subjective than those of color order, thus reversing the order assigned.

Eighth Problem. Quantitative Criteria of the Subjective.—The next problem grows directly out of the preceding one, and has to do with the proposed "quantitative criterion of the subjective." Wells writes: "So far as any distinction on a statistical basis is possible we might consider as subjective those types in which the various judgments of the individual formed a species of their own, varying from each other considerably less than from an equal number of judgments made by different individuals; and consider as objective those in which an individual would vary from his own independent judgments about as much as the variation of an equal number of

judgments by different individuals. . . . The two categories would almost certainly be continuous" (25—512).

A determination of these criteria for materials affording three classes of judgments was the primary purpose of Wells's study. His conclusion may be given in his own words: "It has appeared that in the first class (the highly subjective feeling of preference for different sorts of pictures) the judgments of each individual cluster about a mean which is true for that individual only, and which varies from that of any other individual more than twice as much as its own judgments vary from it; that in the second class, with the colors, the variability of the successive judgments and that of those by different individuals markedly approached each other but still preserved a significant difference; while in the third class, with the weights, we found that there might be even an excess of the individual variability over the 'social.' This comparison seems to afford, to a certain extent, a *quantitative criterion of the subjective*" (25—547).

Further determinations of a somewhat similar sort may be derived from many of my own studies. Instead of using a figure of variability I have employed the coefficients of correlation. The significance should be the same and fewer trials are required to determine the results.

TABLE LXI
COEFFICIENTS OF SUBJECTIVITY

Material	Trait	Obs.	Average Personal Consistency 2 Trials	Average Agree- ment with the Group Av.	Subjectivity Ratio
Faces (photos) ..	Frankness	10	.625	.632	.99
Faces (photos) ..	Intelligence	10	.627	.583	1.07
Faces (photos) ..	Beauty	10	.724	.641	1.13
Handwriting....	Resemblance	9	.789	.644	1.22
Syllables.....	Agreeableness	10	.687	.532	1.29
Syllables.....	Ease	10	.667	.492	1.36
Jokes.....	Funniness	10	.550	.390	1.41
Appeals.....	Persuasiveness	20	.677	.432	1.57
Faces (photos) ..	Attractiveness	10	.806	.466	1.73

Table LXI. gives a series of these determinations. The various materials and traits are arranged in an order of increasing subjectivity as measured by the "subjectivity ratio" (ratio of index of personal consistency to index of group agreement). Judgments of the frankness and intelligence of faces (photographs) are completely objective, that is, a given individual correlates as closely with the average judgment of the group as he does with his own judgment on another occasion. But as one goes on down through the table the

personal consistency coefficients remain fairly constant while the coefficients of group agreement decrease. This gives a larger and larger "subjectivity ratio," until, in judgments of the attractiveness of faces, the personal consistency coefficients are nearly twice as large as those of group agreement.

The use of the coefficients of correlation as criteria of subjectivity in the case of judgments expressed by serial arrangement is much more satisfactory than the relation of the two figures of variability. Fewer trials are required for the determination, and the measures are not complicated by the *end error*, and other factors which tend to disguise the real size of the M.V.'s.

It is probable, however, that the distinction between subjective and objective judgments is at best but an artificial one. The chief difference between the two classes seems to consist in the amount or clearness of the differences present between the various items of the material judged. Judgments of preference will, in the case of a given individual, be expressed as consistently as judgments of weight, duration or intensity, providing the differences are equally perceptible; and judgments of intensity, etc., will vary as much as those of preference if the differences afforded by the material are sufficiently slight. The fact that a so-called objective scale may be applied to the material in the one case and not in the other, is, in the first place, only an extrinsic fact, and in no way conditions the psychological act of judgment. In the second place the objective scale derives its own validity in the long run only from the consensus of opinion and from its pragmatic value. So far as this is concerned a consensus of opinion may be secured for even the most variable and personal sort of material, as witness Thorndike's scales for measuring the excellence of penmanship, literary composition, drawing, etc. The only difference between the two cases would be in the universality of the verdict, and this again in no way conditions the psychological act. It is apparent that the coefficients are merely indices of certain characteristics of the material, rather than of any features of the judgments, as judgments. A certain sort of material may not be constant from time to time or from observer to observer (jokes or comic pictures, for examples). Here the judgment attitude may be conceived as constant, but the material changed. Or one sort of material may provide larger differences between items most alike, and either situation would be revealed by the "coefficients of subjectivity."³ It

³ It is of course also true that, in judging such a general trait as "attractiveness" different observers may proceed on the basis of different qualitative standards and this fact would also be reflected in several of the coefficients, though not in all of them.

is to be expected that various sets of material, of the same content but with differing degrees of difference between successive items would show the same differences in "subjectivity" as those found with different kinds of material. Subjectivity means, then, either of two things, or both: (1) The amount of difference, (2) the universality of the verdict. These also differentiate judgment and perception.

Ninth Problem. Agreement Between Diverse Groups.—The final problem to be presented here concerns the agreement between the average judgments of two groups of observers, when only small groups are used. It is of course obvious that if the two groups are sufficiently large and represent similar or random selections of humanity, the two final orders will be identical, no matter how "subjective" the material may be. But if the groups are small, or if they represent different samplings of human nature, differences might be expected which would be of interest to individual, social, and applied psychology.

I have brought together in the following table such material as I have been able to secure from my own studies and from the published reports of others. The range of material represented is small, and this problem would seem to constitute an interesting theme for further work in statistical psychology.

In the case of this sort of material the average correlation of two groups representing approximately the same sampling of the popula-

TABLE LXII

GROUP AGREEMENTS IN EVALUATION

Material, Trait, and Observers	<i>r</i>	P.E.
H. L. H. Appeals, relative persuasiveness.		
20 women with 10 other women.....	.610	.06
20 women with 20 men.....	.624	.06
10 women with 20 men.....	<u>.598</u>	.06
Average of all three coefficients.....	.611	
E. K. Strong. Advertisements. Persuasiveness.		
15 men with 10 women.....	.53	.07
25 subjects and group of advertising experts.....	.51	.10
25 subjects and manufacturers of the commodity.....	.52	.10
Advertising experts and manufacturers.....	<u>.64</u>	.08
Average of all four coefficients.....	.55	
Kuper. Cosmos prints. Preference.		
100 boys with 100 girls (ages 6.5 to 16.5).....	.24	.06
E. K. Strong. Advertisements. Persuasiveness.		
50 college men and 97 farmers and mechanics.....	-.53	.07
22 college women and 30 college women.....	.93	.02

tion is about .60. The average personal consistency coefficient is about .70, while the correlation of two trials by the same group on two different occasions is about .90. The coefficient of personal consistency thus stands about midway between that of the consistency of a group and the agreement of two diverse groups.

The last two figures from Strong's data, and the one from Kuper's study show the great degree to which the group agreements are conditioned by the composition of the groups. The college students and the manual laborers yield a large negative coefficient, while the two groups of college students give almost perfect positive correlation. The boys and girls correlate, in judging the interest of pictures, by only .24. When college students or adult men and women judge the degree of their interest in appeals not remotely different in character from those used with the children, men and women show as high correlation as do two groups from the same sex. It would seem that in this index of group correlation we have then another useful index of the subjectivity of the material. If the material were weights or brightness intensities there would be no reason for expecting these various groups to show any significant differences in the degree of mutual correlation.

We are thus provided with at least five different indices of subjectivity,—personal consistency, approximation to group average, the ratio of these two indices, the ratio of variability to series length ($M.V./P$), and the agreement of diverse groups. It would be interesting to work out the interrelations of these various indices in different judgment situations.

BIBLIOGRAPHY OF THE ORDER OF MERIT METHOD

1. Barrett, The Order of Merit Method and the Method of Paired Comparisons, *Jour. Phil.*, July 3, 1913, 382-4.
2. Cattell, The Time of Perception as a Measure of Difference in Intensity. *Phil. Stud.*, 1903.
3. Cattell, A Statistical Study of Eminent Men, *Pop. Sci. Mo.*, 53, 357, 1903.
4. Cattell, Statistics of American Psychologists, *Am. J. Psychol.*, 1903, XIV, 310.
5. Cattell, Statistical Study of American Men of Science, *Science*, N. S., XXIV.
6. Cattell, A Further Statistical Study of American Men of Science, *Science*, N. S., XXXII.
7. Cattell, Appendix, American Men of Science, 2d ed., 1910.
8. Downey, Study of Family Resemblance in Handwriting, Bulletin No. 1, Dept. of Psychology, Univ. of Wyoming, 1910.
9. Fernald, G. E., The Defective Delinquent Class, Differentiating Tests, *Amer. Jour. of Insanity*, 69, 125-142, 1912.
10. Hillegas, Milo B., A Scale for the Measurement of Ability in English Compositions, Teachers College Studies.

11. Hollingworth, Judgments of the Comic, *Psych. Rev.*, 1911, 18, 132.
12. Hollingworth, Judgments of Persuasiveness, *Psych. Rev.*, 1911, 18, 234.
13. Hollingworth, Influence of Form and Category, *Jour. Phil.*, 1912, 9, 513.
14. Hollingworth, Principles of Appeal and Response, Appletons, 1913.
15. Hollingworth, Experimental Studies in Judgment, *ARCH. OF PSYCH.*, No. 29.
16. Kuper, Group Differences in the Interests of Children, *Jour. Phil.*, 1912, 9, 376.
17. Norsworthy, Validity of Judgments of Character, Essays in Honor of William James, 1908.
18. Strong, The Relative Merits of Advertisements, *ARCH. OF PSYCH.*, 1911, 17.
19. Strong, Application of the Order of Merit Method to Advertising, *Jour. Phil.*, October 26, 1911, 600-606.
20. Strong, Psychological Methods as Applied in Advertising, *Jour. Ed. Psychol.*, Sept., 1913, 393.
21. Sumner, A Statistical Study of Belief, *Psych. Rev.*, 5, 616.
22. Thorndike, Handwriting, Teachers College Record.
23. Thorndike, Mental and Social Measurements, 2d ed., 1913.
24. Wells, A Statistical Study of Literary Merit, *ARCH. OF PSYCH.*, 1907, 7.
25. Wells, On the Variability of Individual Judgments, Essays in Honor of William James, 1908, 511.
26. Yerkes, Introduction to Psychology, Holt, 1911, Ch. XIV.

2



UNIVERSITY OF CALIFORNIA LIBRARY

Los Angeles

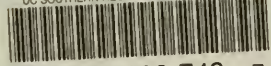
This book is DUE on the last date stamped below.

DEC 30 1967

JUN 12 1968



UC SOUTHERN REGIONAL LIBRARY FACILITY



A 000 289 743 7

